

NASA Contractor Report 201377



National Aeronautics and Space Administration
(NASA)/American Society for Engineering Education
(ASEE) Summer Faculty Fellowship Program - 1995

Volume 1

William A. Hyman and Donn G. Sickorez, Editors

Grant NGT 44-001-800

August 1996

NASA Contractor Report 201377

**National Aeronautics and Space Administration (NASA)/American
Society for Engineering Education (ASEE) Summer Faculty
Fellowship Program - 1995**

Volume 1

William A. Hyman, Editor
*Texas A&M University
College Station, Texas*

Donn G. Sickorez, Editor
*Lyndon B. Johnson Space Center
Houston, Texas*

Grant NGT 44-001-800

August 1996

This publication is available from the NASA Center for Aerospace Information,
800 Elkridge Landing Road, Linthicum Heights, MD 21090-2934, (301) 621-0390

Preface

The 1995 National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship at the Lyndon B. Johnson Space Center (JSC), including the White Sands Test Facility, was conducted by Texas A&M University and JSC. The program at JSC, as well as the programs at other NASA centers, was funded by the Education Division, Higher Education Branch, NASA Headquarters, Washington, D.C., with additional funds from the centers. The objectives of the program, which began nationally in 1964 and at JSC in 1965, are

- To further the professional knowledge of qualified engineering and science faculty members.
- To stimulate an exchange of ideas between participants and NASA.
- To enrich and refresh the research and teaching activities of the participants' institutions.
- To contribute to the research objectives of the NASA centers.

Each faculty participant spent at least 10 weeks at JSC engaged in a research project in collaboration with a NASA/JSC colleague. In addition to the faculty participants, the 1995 program included 5 students. The reports of two of the students are integral with that of the respective fellow. Three students wrote separate reports. Volume 1 contains reports 1 through 15. Volume 2 contains reports 16 through 27 and the three student reports, S-1 through S-3.

Contents

Volume 1

Acharya, R.S.	Characterization of Microgravity Effects on Bone Structure and Strength Using Fractal Analysis.....	1-1
Akl, F.A.	Test Outline for Flutter Analysis of Rectangular Panels in Rarefied Flow Conditions	2-1
Anderson, G.T.	Predictive Modeling of Cardiac Ischemia	3-1
Baker, S.D.	Some Considerations in the Determination of the Accuracy of a Measurement in Space of the Newtonian Gravitational Constant (G).....	4-1
Barrera, E.V. and Olivas, J.D.	Material Assessment of Components of the Extravehicular Mobility Unit.....	5-1
Coombs, C.R.	Regional and Localized Pyroclastic Deposits on the Moon.....	6-1
Corey, K.A.	Design of Plant Gas Exchange Experiments in a Variable Pressure Growth Chamber.....	7-1
Duncan, A.B. and Sciascia, V.M.	Investigation of Two-Phase Flows in Piping Bends and Elbows.....	8-1
Hall, E.B.	A Velocity Distribution Model for Steady State Heat Transfer	9-1
Hogan, H.A.	Trabecular Bone Mechanical Properties and Fractal Dimensions.....	10-1
Janikow, C.Z.	Constraints in Genetic Programming.....	11-1
Khorsandi, M.	Orbiter Flight Deck Redesign (A Physical Layout for a Futuristic Flight Deck).....	12-1
Kleis, S.J.	Fluid Dynamic Verification Experiments on STS-70.....	13-1
Knopp, J.	LCTV Holographic Imaging	14-1
Leon, J.V.	Intelligent Planning and Scheduling for Controlled Life Support Systems	15-1

Contents

Volume 2

Loganantharaj, R.	Theory and Ontology for Sharing Temporal Knowledge ...	16-1
McGinnis, M.R.	Evaluation of Automated Yeast Identification Systems	17-1
Pate, D.W.	A Human Factors Analysis of EVA Time Requirements....	18-1
Perry, E.	Stable Isotope Analysis of Molecular Oxygen From Silicates and Oxides Using CO ₂ Laser Extraction	19-1
Richardson, A.O.	Iliad Testing and a Kalman Filter for 3-D Pose Estimation.....	20-1
Rob, M.A.	Feasibility Study of Velocity and Temperature Measurements of an Arejet Flow Using Laser Resonance Doppler Velocimetric (LRDV) Technique	21-1
Shearrow, C.A.	Design and Implementation of a Windows NT Network to Support CNC Activities.....	22-1
Snow, L.D.	Dextroamphetamine: A Pharmacologic Countermeasure for a Space Motion Sickness and Orthostatic Dysfunction	23-1
Tan, A.	Detecting the Dynamical State of the Atmosphere From the Orbital Decay of the Oederacs Sphere	24-1
Taneja, V.S.	Proportional and Scale Change Models to Project Failures of Mechanical Components With Applications to Space Station	25-1
Watson, J.C.	An Optimal User-Interface for Epims Database Conversions and SSQ 25002 EEE Parts Screening	26-1
Waugaman, D.	Support of Gas Flowmeter Upgrade.....	27-1

Contents

Student Papers

Cauthen, C.	Volcanism in Eastern Africa.....	S-1
Marble, E.	Measuring the Spectral Emissivity of Thermal Protection Materials During Atmospheric Reentry Simulation	S-2
Whitaker, L.O.	An Intelligent Crop Planning Tool for Controlled Ecological Life Support Systems.....	S-3

**Characterization of Microgravity Effects on Bone Structure and
Strength Using Fractal Analysis**

Final Report

NASA/ASEE Summer Faculty Fellowship Program-95

Johnson Space Center

Raj S. Acharya
Associate Professor
Electrical & Computer Engineering
State University of New York at Buffalo
Buffalo, NY 14260
acharya@eng.buffalo.edu

Linda Shackelford
Technical Lead, Bone & Muscle Lab
NASA Johnson Space Center
Houston, Texas 77058

1 Abstract

Protecting humans against extreme environmental conditions requires a thorough understanding of the pathophysiological changes resulting from the exposure to those extreme conditions. Knowledge of the degree of medical risk associated with the exposure is of paramount importance in the design of effective prophylactic and therapeutic measures for space exploration. Major health hazards due to musculoskeletal systems include the signs and symptoms of hypercalciuria, lengthy recovery of lost bone tissue after flight, the possibility of irreversible trabecular bone loss, the possible effect of calcification in the soft tissues, and the possible increase in fracture potential. In this research, we characterize the trabecular structure with the aid of fractal analysis. Our research to relate local trabecular structural information to microgravity conditions is an important initial step in understanding the effect of microgravity and countermeasures on bone condition and strength. The proposed research is also closely linked with Osteoporosis and will benefit the general population.

1 Hypothesis/Rationale

- The rationale for this research is based on the premise that microgravity conditions change bone structure in addition to bone mass.
- Bone structure can be characterized by fractal geometry.
- Fractal characterization of bone structural changes due to microgravity conditions is not only optimal but also pragmatic.

2 Specific Aims

The overall goal is the characterization of bone structural changes due to microgravity with the aid of fractals.

- We propose the use of the Alternating Sequential Filters (ASF) method to estimate the fractal dimension of images. When only small window sizes are available, Continuous Alternating Sequential Filters (CASF) will be used for fractal dimension estimation.
- We compute the fractal dimension of subjects participating in the bed rest study.
- We apply fractal analysis to samples of human bone and relate it to mechanical strength.

3 Significance

Protecting humans against extreme environmental conditions requires a thorough understanding of the pathophysiological changes resulting from the exposure to those extreme conditions. Knowledge of the degree of medical risk associated with the exposure is of paramount importance in the design of effective prophylactic and therapeutic measures for space exploration. Major health hazards due to musculoskeletal systems include the signs and symptoms of hypercalciuria, lengthy recovery of lost bone tissue after flight, the possibility of irreversible trabecular bone loss, the possible effect of calcification in the soft tissues, and the possible increase in fracture potential. Our research to relate local trabecular structural information to microgravity conditions is an important initial step in understanding the effect of microgravity and countermeasures on bone condition and strength. The proposed research is also closely linked with Osteoporosis and will benefit the general population.

4 Background

The effect of micro-gravity on the musculoskeletal system is currently being studied. Significant changes in bone and muscle have been shown after long term space flight. Similar changes have been demonstrated due to bed rest. Bone demineralization is particularly profound in weight bearing bones. Much of the current techniques to monitor bone condition use bone mass measurements. However bone mass measurements do not completely describe the mechanisms to distinguish Osteoporotic and Normal subjects.¹ It has been shown that the overlap between normals and osteoporosis is found for all of the bone mass measurement technologies: single and dual photon absorptiometry, quantitative computed tomography and direct measurement of bone area/volume on biopsy as well as radiogrammetry. A similar discordance is noted in the fact that it has not been regularly possible to find the expected correlation between severity of osteoporosis and degree of bone loss .

Structural parameters such as trabecular connectivity have been proposed as features for assessing bone conditions.¹ It has been shown that in vertebral crush fracture patients, elements such as vertical trabeculae are retained more or less intact, while elements such as horizontal bracing trabeculae are resorbed entirely⁵⁶ . This results in disconnection of large number of trabecular elements. However, in non-fracture patients connections between elements were preserved. Long vertical trabeculae are subject to buckling under loading. When they lose their lateral connections to adjacent trabeculae, the degree of buckling may exceed the inherent strength of the bone. *Structure can be thus be seen as an important feature in assessing bone condition. In this research, we propose the use of fractals to model the trabecular bone structure.*

5 Fractal Analysis

In the past few years a significant amount of effort has been devoted to the study of chaotic phenomena. A part of this effort is directed towards the study of fractals. As defined by Mandelbrot,¹² fractals are surfaces whose dimensions are strictly greater than their topological dimensions. Intuitively, fractals are surfaces which are embedded in between the m and $m + 1$ dimensional manifolds. A fractal object can be characterized by it's dimension which may be interpreted as the quantity of space occupied by the object between the m and $m + 1$ dimensional manifolds. Various alternative definitions of dimension have been proposed in the literature. For a detailed discussion of these different definitions the reader is referred to.¹⁴¹⁵¹⁶ This research employs the definition commonly know as the capacity dimension.¹⁵

It was Kolmogorov³¹ who originally proposed the capacity of a set as

$$d = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)} \quad (1)$$

where if the set in question is a bounded subset of a p -dimensional Euclidean space R^p , then $N(\epsilon)$ is the minimum number of p -dimensional cubes of side ϵ needed to cover the set. *For a Fractal set, d can be a non-integer.*

Fractals are known to obey the self-similarity property. Self-similarity implies that the statistical properties of a fractal are independent of the scale or resolution of observation.

There are number of ways in which fractal geometry can be applied to the analysis of image texture. Pentland et al¹¹ used a method related to the co-occurrence matrix technique of texture classification based on fractal dimension. They found the standard deviations of the difference of gray levels separated by a given vector and plotted it against the vector lengths as a log-log graph. Maragos has used Morphological Covers to estimate Fractal dimension of Synthetic Images.¹⁷ In another technique⁷ a two dimensional gray level image is represented as a three-dimensional surface whose height at each point represents the gray level at that point and the surface area is measured at different scales. It has also been shown that⁸ an n -dimensional fractal object can be characterized by the fractional Brownian motion of n variables and that the relationship between the power spectral density and r are independent of the projection.³⁰ This makes the fractal dimension computed from the projections of an n -dimensional fractal object represent the original object. Several studies have modeled radiographic bone images using fractional Brownian motion and have used maximum likelihood estimation to find the fractal dimension.^{8,9}

A recent study has indicated that fractal analysis can distinguish between dental radiographs of pre and post menopausal women.¹³ Analyses of dental radiographs have been shown to be independent of imaging conditions involved in taking the radiographs (such as the angle between x-ray collimator and anatomical structure of interest) to a significant extent.¹⁰ These preliminary studies also suggest that the fractal measure is to a major extent, independent of the anatomical site being analyzed. Another study used fractal dimensions to attempt predicting osseous changes in ankle fractures.²⁹

A variety of methods have been proposed to estimate the fractal dimension of trabecular bone structures.^{18,19,22,20} *An important problem is the estimation of fractal dimension from images when only small window sizes of the desired structures are available.* In this proposal, a new method based on Alternating Sequential Filters (ASF), is presented to estimate the dimension of a fractal object. When only small window sizes (between 16 x 16 to 64 x 64) are available, we propose the use of Continuous Alternating Sequential Filters (CASF) for fractal dimension estimation. Frequently MRI studies are carried out to provide images of the spine. The individual vertebra occupies only a small region of the entire image of the spine (window size less than 50 X 50). CASF methods are well suited to handle these small window size situations. However, if the window size is greater then 64 x 64, ASF methods can be used.

5.1 Alternating Sequential Filter (ASF)

The method presented here to estimate fractal dimension is based on using Alternating Sequential Filters (ASF). Due to space limitation, a very brief review of ASF is presented. For further details on the morphological concepts presented here the reader is referred to.⁴²¹²²²⁶ Alternating Sequential Filters appeared initially in the work of Sternberg.²⁷ The image X is filtered by a closing operation ϕ_1 followed by an opening operation γ_1 , then it is filtered again by a larger closing ϕ_2 and a larger opening γ_2 , etc..., which in essence produces the operator

$$M_i(X) = \phi_i \gamma_i \dots \phi_1 \gamma_1(X) \quad (2)$$

Transformations that apply products of openings and closing in general introduce less distortions than individual operation such as dilations. The operations openings and closing, tend to preserve the 'rough' nature of the image X . Although the use of dilations and erosions have been proposed in the literature to estimate the fractal dimension,²²² it was observed in the scope of this work that ASF's are more suitable for estimating fractal dimension.

Let \mathcal{L} be a complete lattice. Now define two mappings, a pair of primitives, $(\lambda, X) \rightarrow \gamma_\lambda(X)$ and $(\lambda, X) \rightarrow \phi_\lambda(X)$ from $\mathbb{R}^+ \times \mathcal{L}$ into \mathcal{L} such that for all $\lambda > 0$, γ_λ is an opening and ϕ_λ is a closing such that

$$\lambda \geq \mu \Rightarrow \gamma_\lambda < \gamma_\mu \text{ and } \phi_\lambda > \phi_\mu \quad (3)$$

Note that λ represents the size of the structuring elements used and X is any arbitrary set on \mathcal{L} . Now let m_λ be an operator defined as:

$$m_\lambda = \gamma_\lambda \phi_\lambda \quad (4)$$

and for pairs $\lambda, \lambda' \in \mathbb{R}^+$ with $\lambda' > \lambda$ construct the sequence of products

$$\begin{aligned} M_0 &= M_0(\lambda, \lambda') = m_\lambda m_{\lambda'} , \\ M_1 &= M_1(\lambda, \lambda') = m_\lambda m_{(\lambda+\lambda')/2} m_{\lambda'} , \\ &\vdots \\ M_k &= M_k(\lambda, \lambda') = m_\lambda \dots m_{\lambda+i2^{-k}(\lambda'-\lambda)} \dots m_{\lambda+2^{-k}(\lambda'-\lambda)} m_{\lambda'} , 0 \leq i \leq 2^k \end{aligned} \quad (5)$$

A morphological filter called an Alternating Sequential Filter with primitives γ and ϕ and bounds λ and λ' is defined to be:

$$M = M_\lambda^{\lambda'} = \wedge M_k(\lambda, \lambda'), \quad (6)$$

where M is the infimum over all M_k . The mapping M is increasing and idempotent for $\lambda' > \lambda$.

Once an ASF representation of the image is obtained, the fractal dimension can be easily computed with the aid of equation 1.

5.2 Continuous Alternating Sequential Filters (CASF)

We propose the use of CASF when only small window sizes of the desired structure are available in the image. With larger window sizes, ASF methods can be used. The lattice \mathcal{L} is assumed to be a locally compact Hausdorff topological space. In the digital case the definition of compactness of a space has to be handled more carefully. Let \mathcal{G} be the sampled version of an arbitrary d -dimensional function \mathcal{F} , where $m < d$ is the topological dimension of \mathcal{F} . It is assumed that \mathcal{G} is the only function available. All characteristics of \mathcal{F} at higher resolutions are lost and as such have to be estimated from \mathcal{G} . Compactness of the function \mathcal{G} has to be considered at each scale value at which the function \mathcal{G} is being observed. The following assumptions are made about an arbitrary function \mathcal{F} :

\mathcal{F} is a measurable, continuous, convex and a smooth function with known number of continuous derivatives.

A variety of techniques can be employed to reconstruct the function \mathcal{F} from the function \mathcal{G} given that some of the above assumptions are satisfied.²⁸

Let \mathcal{G} be a d -dimensional fractal set. It is noted that d is not necessarily an integer. Also \mathcal{G} is embedded between m and $m + 1$ dimensional euclidean manifolds such that $m < d < m + 1$. Let $\overline{\mathcal{G}} = \text{ASF}(\mathcal{G})$ be a morphological mapping from $\mathbb{R}^+ \times \mathcal{L}$ into \mathcal{L} over some arbitrary neighborhood U defined by the size of the structuring element \mathbf{B} . Note that the ASF is being used as the locally smoothing operator M described above. The use of ASF filters preserves the global nonlinear characteristics of \mathcal{G} while performing some level of smoothing over the local neighborhood U . Let $\mathcal{G}_\varepsilon = T(\overline{\mathcal{G}})$ be the reconstructed object \mathcal{G} at a resolution scale ε and $\overline{\mathcal{G}}_\varepsilon = \text{ASF}(\mathcal{G}_\varepsilon)$. The sets $\overline{\mathcal{G}}_\varepsilon$ provide a pyramidal representation as different scale values ε are used and the size of $\overline{\mathcal{G}}_\varepsilon$ is changed accordingly. As in (1), let $N(\varepsilon, U)$ be the number of m -dimensional cubes required to cover the neighborhood U of $\overline{\mathcal{G}}_\varepsilon$. Over any arbitrary neighborhood U , the function $N(\varepsilon, U)$ is a mapping from $\mathbb{R}^+ \times \mathcal{L} \rightarrow \mathbb{R}^+$. The fractal dimension d can then be estimated by:

$$\log \sum_U N(\varepsilon, U) = d \log \varepsilon \quad (7)$$

Note that the sum above is taken over all compact, closed, neighborhoods U of $\overline{\mathcal{G}}_\varepsilon$.

6 Preliminary Experimental Results.

6.1 Bedrest Studies

Fractal analysis was performed on calcaneus regions of subjects undergoing bedrest study. Fig 1 shows one slice of the calcaneus region imaged with the MRI scanner. Figure 2 shows the

location of 5 different window locations in which the fractal dimension was computed. Figure 3 shows that the fractal dimension of the calcaneus region as a function of time.

6.2 Fractal Analysis of Tibia

MRI images of an isolated tibia were obtained. Fig 4 shows a slice of the MRI image of the tibia. Fig 5 shows the location of the 53 windows in which the fractal dimension was computed. Fig 6 shows the plot of the fractal dimension as a function for the 53 different window locations in the image.

7 Summary and Conclusions

We have used the fractal dimension to characterize the trabecular bone structure. We have computed the fractal dimension in the calcaneus region of a bedrest subject. We have also computed the fractal dimension of an isolated tibia scanned in the MRI scanner. We plan on relating the fractal dimension to mechanical strength in an ongoing research project.

8 REFERENCES

- [1] R. P. Heaney, 'Osteoporotic Fracture Space: An Hypothesis', Bone and Mineral, vol6, pp1-13, 1989.
- [2] J. Samarabandu, R. Acharya, E. Hausmann, and K. Allen, "Analysis of Bone X-Rays Using Morphological Fractals", *IEEE Trans. on Medical Imag.*, Vol. 12., No. 3, Sept. 1993.
- [3] N. Gache et al, "Fractal Dimension Estimators for Fractional Brownian Motions" pp3557-3560, ICASSP 91.
- [4] A. Morales and R. Acharya, "An Image Pyramid with Alternating Sequential Filters", *IEEE Transactions on Image Processing*, June 1995. (In Press)
- [5] E. Kleerekoper et al, "Calcaneus Bone Architecture and Bone Strength", in Editor: Christensen C, Osteoporosis 1987, Viborg, Denmark: Norhaven A/s-1987, pp 289-300.
- [6] J. E. Aaron et al, "The Microanatomy of Trabecular Bone Loss in Normal Aging Men and Women", *Clin Orthop Res*, vol 215, pp260-271, 1987.
- [7] S. Peleg et al, "Multiple Resolution Texture Analysis and Classification", *IEEE PAMI*, vol 6, No 4, pp518-523, July 1984.

- [8] T. Lundhal et al, "Fractional Brownian Motion: A Maximum Likelihood Estimator and It's Application to Image Texture", *IEEE TMI*, vol 5, pp152-161, Sep 86.
- [9] C. Fortin et al, "Fractal Dimension in the Analysis of Medical Images", *IEEE Engineering in Medicine and Biology Magazine*, vol 11, pp65-71, 1992.
- [10] R. L. Webber et al, "Evaluation of Site Specific Differences in Trabecular Bone Using Fractal Geometry", *Journal of Dental Research*, Abstract no 2095, 1991.
- [11] A. Pentland, "Fractal Based Description of Natural Scenes", *IEEE-CVPR*, pp209-210, June 1983.
- [12] B. B. Mandelbrot, "Fractals: Form, Chance and Dimension", Freeman, San Francisco, 1977.
- [13] U. E. Ruttimann, "The Use of Fractal Geometry to Quantitate Bone Structures from Radiographs", *J Dental Research*, Abstract 1431, 1990.
- [14] J. D. Farmer, E. Ott, J. A. Yorke, "The Dimension of Chaotic Attractors", *Physica 7D*, pg. 153 -180, North-Holland Publishing Company, 1983.
- [15] T. S. Parker and L. O. Chua, *Practical Numerical Algorithms for Chaotic Systems*, pg. 167-199, Springer-Verlag 1989.
- [16] A. Lasota and M. C. Mackey, *Chaos, Fractals and Noise*, pg. 432-437, Second Edition, Springer-Verlag, 1994.
- [17] P. Maragos et al, "Measuring Fractal Dimension of Signals: Morphological Covers and Iterative Optimization", pp 108-121, *IEEE Transactions on Signal Processing*, vol 41, No 1, Jan 1993.
- [18] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pg. 674-693, Jul. 1989.
- [19] U. E. Ruttiman and J. A. Ship, "The Use of Fractal Geometry to Quantitate Bone Structure From Radiographs", *J. Dent. Res.*, pg. 69, 1990. Abstract no. 1431.
- [20] S. Majumdar et al, "Application of Fractal Geometry Techniques to The Study of Trabecular Bone", *Med. Phys.* 20 (6), Nov/Dec 1993.
- [21] J. Serra, *Image Analysis and Mathematical Morphology*, New York, Academic Press, 1982.
- [22] J. Serra, *Image Analysis and Mathematical Morphology*, Vol. 2, New York, Academic Press, 1988.
- [23] C. L. Benhamou et al, "Fractal Organization of Trabecular Bone Images on Calcaneus Radiographs", pp1909-1918, *Journal of Bone and Mineral Research*, vol 9, No 12, 1994.

- [24] Von Mises, Richard, "Mathematical Theory of Probability and Statistics", Academic Press, 1964.
- [25] SPSS: Statistical Package for the Social Sciences, ed: Norman H. Niles, McGraw Hill.
- [26] C. Giardina and E. Dougherty, *Morphological Methods in Image and Signal Processing*, New Jersey, Prentice Hall 1988.
- [27] S. R. Sternberg, "Morphology for Grey Tone Functions", *Computer Vision Graphics and Image Processing*, Vol. 35, 1986.
- [28] A. P. Korostelev and A. B. Tsybakov, *Minimax Theory of Image Reconstruction*, Springer-Verlag, 1993.
- [29] R. L. Webber et al, "Predicting Osseous Changes in Ankle Fractures", *IEEE Engineering in Medicine and Biology*, pp103-110, March 1993.
- [30] W. S. Kulinski et al, "Application of Fractal Texture Analysis to Segmentation of Dental Radiographs", vol 1092, *Medical Imaging-3, SPIE Proceedings*, 1989.
- [31] A. N. Kolmogorov, "A New Invariant for Transitive Dynamical Systems", *Dokl. Acad. Nauk SSSR*, pg. 861-864, No. 119, 1958

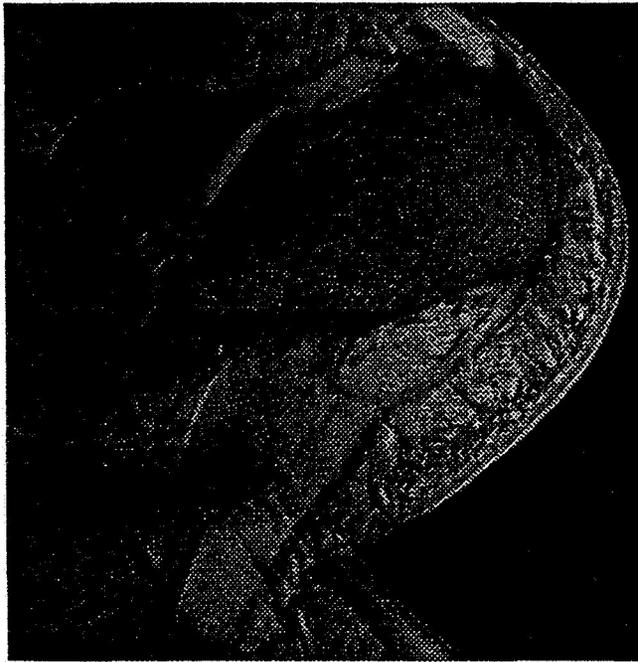


Fig 1

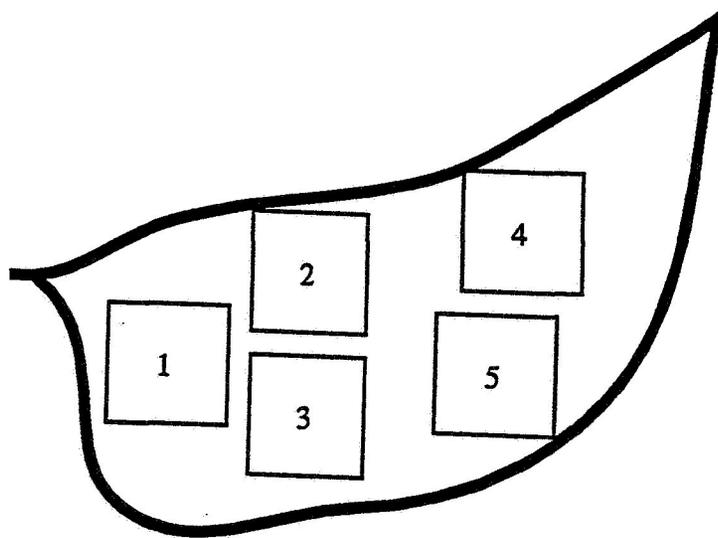


Fig 2

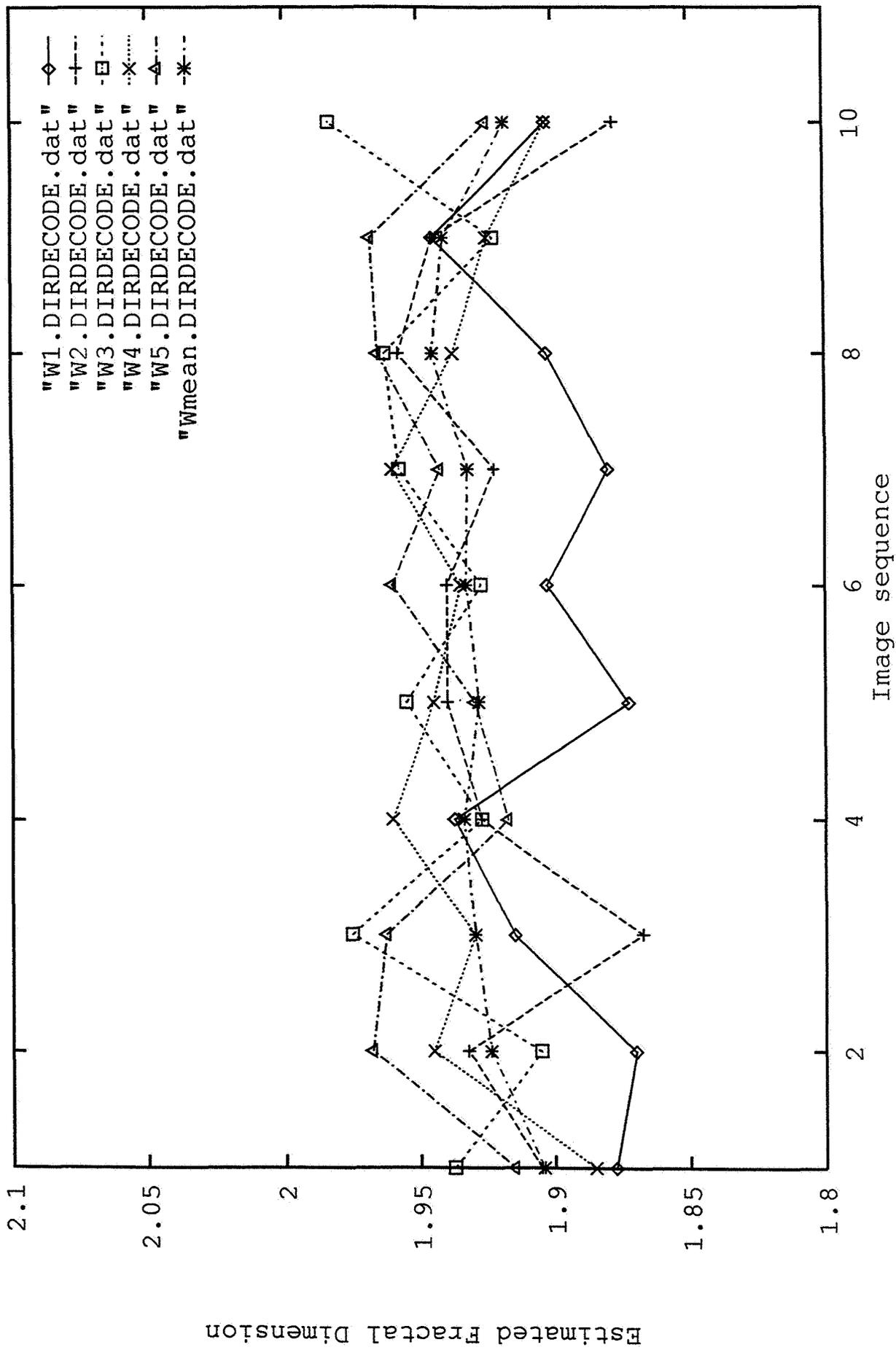


Fig 3

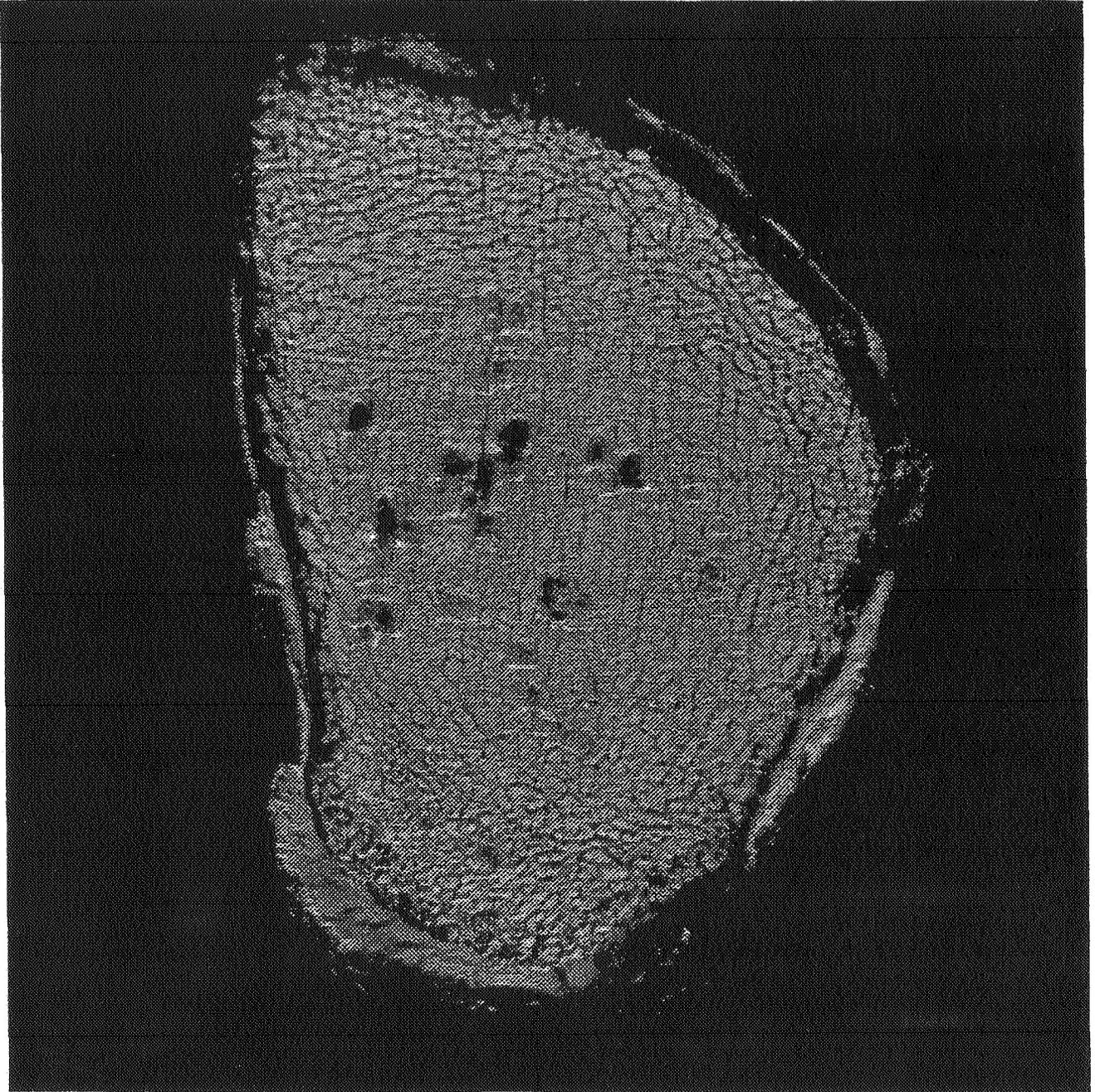


Fig 4

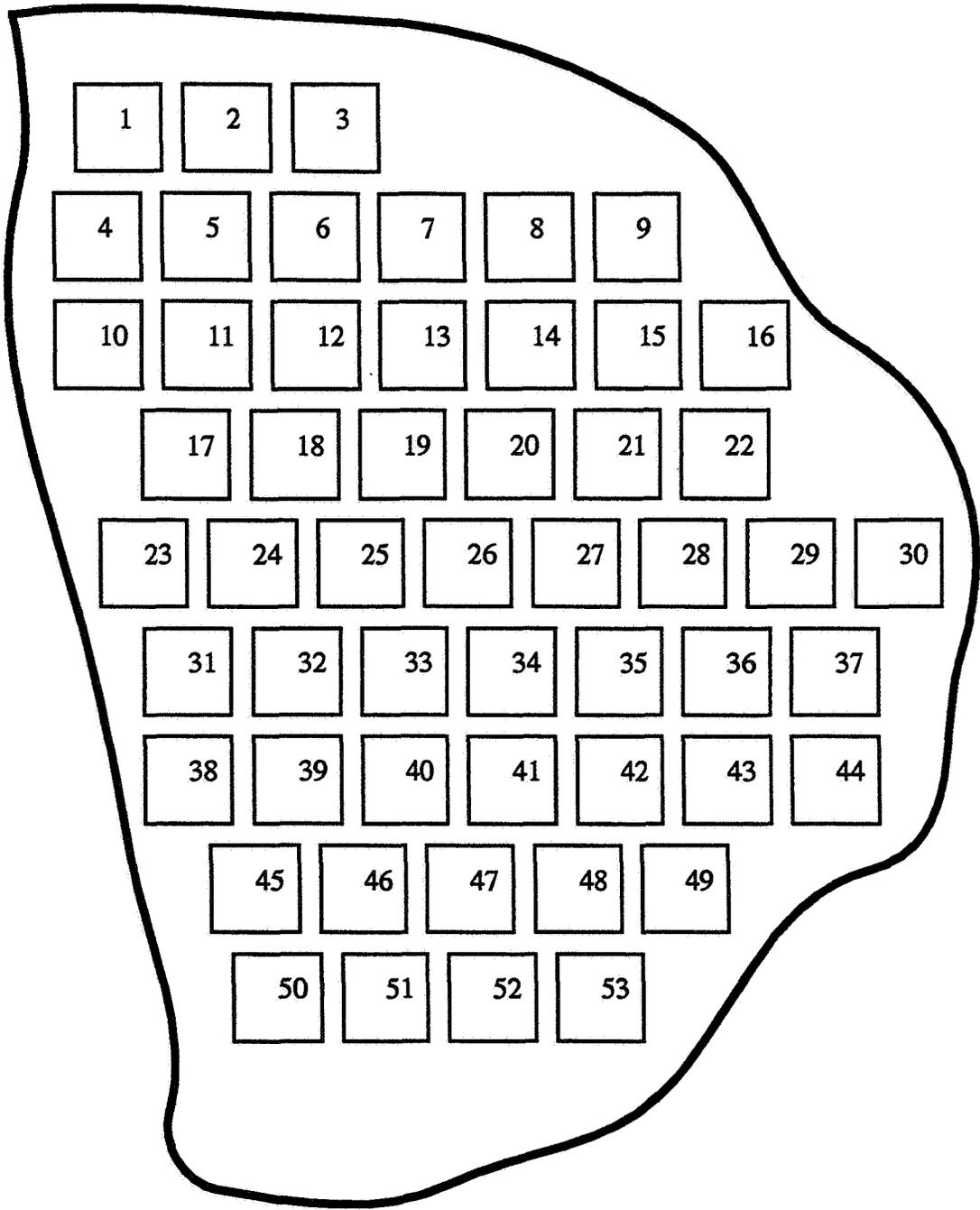
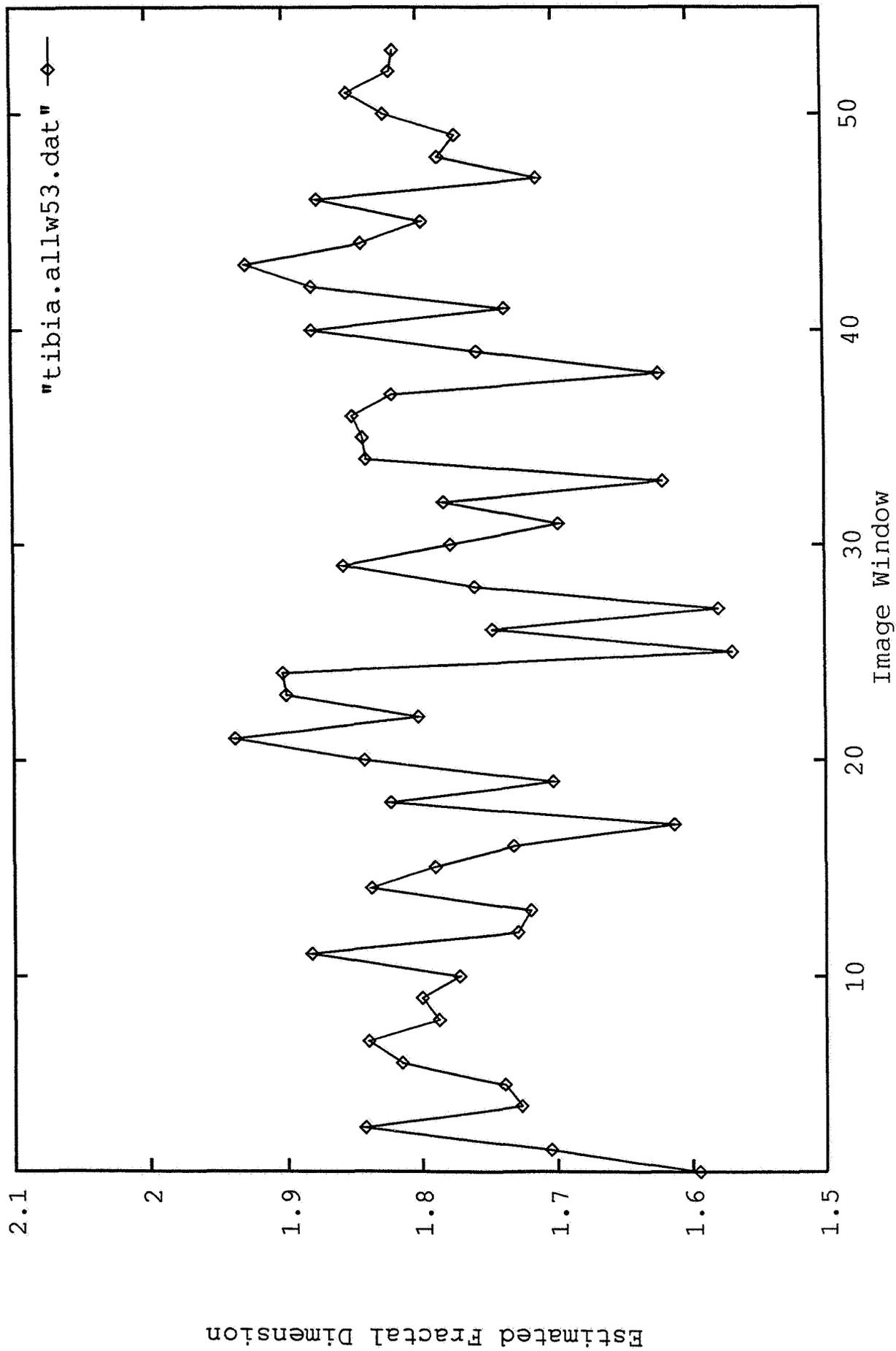


Fig 5

Fig 6



**TEST OUTLINE FOR FLUTTER ANALYSIS OF
RECTANGULAR PANELS IN RAREFIED FLOW CONDITIONS**

Final Report
NASA/ASEE Summer Faculty Fellowship Program--1995
Johnson Space Center

Prepared By:	Fred A. Akl, Ph.D., P.E.
Academic rank:	Professor
University & Department:	Louisiana Tech University Department of Civil Engineering Ruston, Louisiana 71272
 NASA/JSC	
Directorate:	Engineering
Division:	Structures and Mechanics
Branch:	Structures and Dynamics
JSC Colleague:	James D. Dagen
Date Submitted:	August 7, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

Jet plume impingement forces acting on large flexible space structures may precipitate dynamically unstable behavior during space flights. Typical operating conditions in space involve rarefied gas flow regimes which are intrinsically distinct from continuum gas flow and are normally modeled using the kinetic theory of gas flow. Docking and undocking operations of the Space Shuttle with the Russian Mir space laboratory represent a scenario in which the stability boundaries of solar panels may be of interest. Extensive literature review of research work on the dynamic stability of rectangular panels in rarefied gas flow conditions indicated the lack of published reports dealing with this phenomenon.

A recently completed preliminary study for NASA JSC dealing with the mathematical analysis of the stability of two-degree-of-freedom elastically-supported rigid panels under the effect of rarefied gas flow was reviewed. A test plan outline is prepared for the purpose of conducting a series of experiments on four rectangular rigid test articles in a vacuum chamber under the effect of continuous and pulsating Nitrogen jet plumes. The purpose of the test plan is to gather enough data related to a number of key parameters to allow the validation of the two-degree-of-freedom mathematical model. The hardware required careful design to select a very lightweight material while satisfying rigidity and frequency requirements within the constraints of the test environment. The data to be obtained from the vacuum chamber tests can be compared with the predicted behavior of the theoretical two-degree-of-freedom model. Using the data obtained in this study, further research can identify the limitations of the mathematical model. In addition modifications to the mathematical model can be made, if warranted, to accurately predict the behavior of rigid panels under rarefied gas flow regimes.

INTRODUCTION

This study is a step toward the development of an engineering tool capable of predicting the stability boundaries of solar panels while taking into consideration their unique structural characteristics. Although research on flutter analysis of structures in continuum gas flow regimes is well documented in the published literature, work on flutter analysis of structures in rarefied gas flow is virtually non-existent. In recent years, scientists and engineers at NASA JSC and elsewhere conducted a number of investigations to study the physics of jet plumes and their impingement loads on stationary rigid objects in vacuum chambers and in space (1, 2, 3, 4, 5).

Literature review of the jet plume impingement forces and the dynamic stability of rectangular panels in rarefied gas flow conditions was undertaken. The review focused on a recently completed study dealing with the theoretical analysis of the flutter of a two degree-of-freedom rigid body model (6,7). A brief test outline is given for the purpose of investigating the experimental behavior of rectangular panels in a vacuum chamber under the effect of continuous and pulsating Nitrogen jet plume. Three of the four test articles will be attached to a flexible element which will allow torsional and bending vibrations, and one test article is stationary. Validation of the stability boundaries predicted by the theoretical rigid body model is the primary objective of the test plan.

TEST OVERVIEW

Tests on a total of four rectangular panels of 40" by 20" are outlined. The first test article (TA-I) consists of a rigid metal panel mounted on a rigid support. The second and third test articles (TA-II and TA-III) are elastically-supported rigid panels. The fourth test article (TA-IV) is an elastically-supported flexible panel. The tests will be conducted in Vacuum Chamber B at NASA JSC to study the dynamic stability of rectangular panels when subjected to rarefied gas continuous and pulsating flow in vacuum conditions. A number of key parameters will be varied to study the phenomenon of dynamic stability of these panels under different conditions.

TEST SETUP

The test setup is referenced to a global set of axes (XYZ) whose origin (0,0,0) is located at the center of the floor of Vacuum Chamber B. The XY plane is parallel to the floor. The X-axis is oriented to the geographic east and the Y-axis is oriented toward the geographic north. The Z-axis is vertically upward for positive direction.

The gas nozzle will be located at (0, 0, 10 ft). The longitudinal axis of the nozzle will be parallel to the X-Y plane. The nozzle will be mounted on a turning table which will allow orientation of the nozzle axis to be at an angle θ of 0° for TA-I, 90° for TA-II, 180° for TA-III and 270° for TA-IV, where the angle θ is measured positive counterclockwise with respect to the positive X-axis.

All the test articles (rigid or flexible) will be oriented with their transverse axis parallel to the Z-axis. The orientation of the longitudinal axis of the test articles (angle of attack) will vary during the tests. The longitudinal axis of the jet nozzle will be passing through the geometric center of the test articles during testing. Test article TA-I will be stationary during the tests. Test articles TA-II, TA-III and TA-IV will be supported along the transverse axis to provide the specified torsional frequency of 1 Hz for the test articles/support system while restricting the bending frequency of the system to at least 10 Hz.

The far end of the support for test articles TA-II, TA-III and TA-IV will be secured to a mounting rig at or near the Vacuum Chamber floor such that translations and rotations about the X, Y, and Z axes are restrained. The mounting rig for TA-II shall be attached to a slider which will allow the distance between the test article and the nozzle exit to be varied within ± 1 ft. The mounting rigs for all four test articles will also provide the facility to change the angle of attack of the test articles by rotating them about the Z-axis.

A video camera, laser targets and a number of transducers will be set up on and around the test article. Data acquisition will be setup outside the Chamber for data A/D conversion and collection.

JET PLUME

Type:	Nitrogen
Nozzle:	Dia. = 0.0065 in \pm 0.0005 in single orifice.
Flow:	Continuous until plume collapse
Pulsating flow:	two scenarios: periodic square and periodic half-sine waves. Pulse duration $t_p = 0.1$ sec Pulse separation $\Delta t = 0.2$ sec.

TEST ARTICLES

Test Article I (TA-I)

TA-1 consists of a 40" by 20" rigid aluminum panel of uniform thickness with a Kapton thin film bonded to the front side. The panel will be supported by brackets, or other type of support, which will restrain the plate during the tests. The thickness of the aluminum plate is governed by the distance needed to accommodate the pressure transducers to be attached to the test article. Three pressure transducers will be used to measure the dynamic pressure. Six thermocouples will be used to measure the temperature of the plate surface.

Test Article II (TA-II)

TA-II consists of a 40" by 20" panel. The panel is made of a framework of 1/4" composite tubes. A three mill Kapton film will cover the front side of the panel. A pretension force will be applied to the Kapton film during fabrication. The Kapton film is then secured to the back side of the outer composite tubes. The test article will be mounted on a slider which will allow the distance between the nozzle exit and the front of TA-II to vary from 3 ft to 5 ft. The mass moment of inertia of the test article/support system about the Z-axis, I , is equal to 60 lbm.in². The torsional natural frequency of the test article/support system, f_z , is equal to 1 Hz.

Test Article III (TA-III)

TA-III also consists of a 40" by 20" panel to be made of a honeycomb paper with two triangular cutout areas. A two mill Kapton (or mylar) film covers the front and back sides of the panel. An adhesive film of 5.5 mil will be used to glue the films to the panel. The mass moment of inertia of the test article/support system about the Z-axis, I , is equal to 60 lbm.in². The torsional natural frequency of the test article/support system, f_z , is equal to 1 Hz.

Test Article IV (TA-IV)

TA-IV consists of a 40" by 20" flexible metal panel such that the bending frequency of the panel itself about its transverse axis (Z-axis) is 2 Hz. The mass moment of inertia about the Z-axis, I , is equal to 60 lbm.in². The torsional natural frequency of the test article/support system, f_z , is equal to 1 Hz.

REFERENCES

- [1] Fitzgerald, S.M., Bouslog, S.A. and Hughes, J.R., "Model for Predicting Orbiter PRCS Plume Impingement Loads," NASA JSC Report #26507, Rev. A, June 1995.
- [2] Bouslog, S.A., Fitzgerald, S.M. and Machin, R.A., "Instrument Characterization Tests in a Nitrogen Free Jet," NASA JSC Report #26706, March 1995.
- [3] Fitzgerald, S.M., and Hughes, J.R., "Orbiter PRCS Plume Impingement Toolkit," NASA JSC Report #26583, March 1995.
- [4] Fitzgerald, S.M., Bouslog, S.A. and Hughes, J.R., "Model for Predicting Orbiter PRCS Plume Impingement Loads," NASA JSC Report #26507, January 1994.
- [5] Machin, R.A., et al, "Shuttle Plume Impingement Flight Experiment (SPIFEX) Data Reduction," NASA JSC Report #27030, Vol. I.
- [6] Storch, J.A., "Flutter of Solar Arrays Subjected to Rarefied Plume Impingement," Draper Laboratory Reports, September 1994 and January 1995.
- [7] Storch, J.A., "Solar Array Flutter in Pulsed Rarefied Flow," Draper Laboratory Report, June 1995.

PREDICTIVE MODELING OF CARDIAC ISCHEMIA

**Final Report
NASA/ASEE Summer Faculty Fellowship Program-1995
Johnson Space Center**

Prepared By:	Gary T. Anderson, Ph.D.
Academic Rank:	Associate Professor
University & Department	University of Arkansas at Little Rock Dept. of Applied Sciences Little Rock, AR 72204
NASA/JSC	
Directorate:	Information Systems
Division:	Technology Systems
Branch:	Client/Server
Date Submitted:	8/11/95
Contract Number:	NGT-44-001-800

ABSTRACT

The goal of the Contextual Alarms Management System (CALMS) project is to develop sophisticated models to predict the onset of clinical cardiac ischemia before it occurs. The system will continuously monitor cardiac patients and set off an alarm when they appear about to suffer an ischemic episode. The models take as inputs information from patient history and combine it with continuously updated information extracted from blood pressure, oxygen saturation and ECG lines. Expert system, statistical, neural network and rough set methodologies are then used to forecast the onset of clinical ischemia before it transpires, thus allowing early intervention aimed at preventing morbid complications from occurring. The models will differ from previous attempts by including combinations of continuous and discrete inputs.

A commercial medical instrumentation and software company has invested funds in the project with a goal of commercialization of the technology. The end product will be a system that analyzes physiologic parameters and produces an alarm when myocardial ischemia is present. If proven feasible, a CALMS-based system will be added to existing heart monitoring hardware.

INTRODUCTION

Cardiovascular disease is the leading cause of death in the US, causing about 43% of all mortalities. Each year, more than 5 million patients arrive at Emergency Rooms (ER) with chest pain, with 35-40% of these suffering from acute ischemia [Selker, 1989]. Coronary Care Units (CCUs) have proven to be extremely effective in preventing death from ischemic cardiac events, but the cost of these units limits their presence to only 22% of hospitals. When cardiac patients arrive at a medical facility, a decision must be made as to whether they belong in the CCU or in a less expensive facility such as a Monitored Care Unit (MCU). For patients arriving at a hospital without a CCU, a decision must be made as to whether they can be treated in-house, or should be transported to a tertiary care facility with a CCU.

The cost of wrong triage decisions can be staggering. Estimates of the percentage of patients needlessly admitted to the CCU range from 50% [Rollag, 1992] to 70% [Fineberg, 1984]. Selker [1989] concludes that each year perhaps \$4 billion dollars are spent on CCU care for such patients. In addition, many patients who would benefit from CCU services are not admitted. It is estimated that about 11% [Fleming, 1991] of ER patients with acute ischemic disease are inadvertently sent home. Of those admitted, 9 to 12% [Rollag, 1992; Fleming, 1991] who should be admitted to the CCU are sent to the ward or a step down care facility.

Criteria for admission to a CCU can vary, depending on hospital practice [Weingarten, 1993]. It is known that CCU interventions can significantly lower mortality of patients with acute myocardial infarctions. If implemented in the first 6 - 12 hours after an MI, arrhythmia prophylaxis, cardiac monitoring, thrombolytic therapy and resuscitative interventions available in the CCU can all reduce mortality and morbidity rates for cardiac patients. Quick diagnosis and triage decisions are critical for preventing or effectively treating complications of an MI. However, cardiac triage decisions in the emergency room are often made under severe time pressure, making optimal placements difficult. The proposed CALMS technology will assist the ER physician in making difficult triage decisions by giving them an objective, computer-based second opinion on patient prognosis.

The most difficult triage decision concerns patients with unstable angina, chest pain that is non-responsive to drug treatment. 80-90% of these people will respond to medical therapy, while 10-20% will progress to a myocardial infarction (MI). Based on a pilot study of patients at the University of Arkansas for Medical Sciences, about 8% of people in an MCU will later be transferred to the CCU, indicating that the severity of their illness was originally misinterpreted by the attending cardiologist. Emergency room physicians and family practitioners in rural settings could be expected to have a higher misdiagnosis rate. Once in a CCU, very few life-threatening incidents transpire. If surgery patients, catheterization patients, people admitted to the CCU because they are in the midst of a potentially lethal event and co-morbidity patients (who experience chest pain along with another unrelated illness) are excluded, less than 10% of the remaining population will experience life-threatening episodes. One reason for the low event rate is because of interventions available only in a CCU (e.g., administration of intravenous nitroglycerine or dobutamine), which probably prevented morbid incidences that would have occurred otherwise. However, overcautious admission of people to the CCU likely accounts for a large portion of the low event rate [Selker, 1989].

PREDICTIVE MODELS

Predictive models generally depend on information from a patient's medical history and present medical condition. Several physiologic parameters have been shown to be indicators of future cardiac events. For example, factors as varied as age, hypertension, diabetes, length of stay in CCU [Gheorghide, 1987], ST and T wave changes [Severi, 1988; Bell, 1990], sex, anterior infarction, hypotension at admission, ventricular tachyarrhythmias, diabetes, Killip class III and IV [De Martini, 1990], previous myocardial infarction [Nishi, 1992], and serum urea [Marik, 1990] have all been shown to have short-term prognostic significance. Recently, changes in heart rate variability has also been shown to be a precursor of clinical ischemia [Bianchi, 1993].

Several researchers have developed models to predict which patients could most benefit being in the CCU [Pozen, 1984; Brush, 1985; Weingarten, 1989, Selker, 1991]. Pozen *et. al* developed a model based on seven discrete inputs to the logistic equation. This model worked best at excluding patients from the CCU (rather than predicting who should be admitted), but missed some obvious candidates [Green, 1988]. In addition, two of the criteria can not be reliably found in a patients medical records (nitroglycerine use and history of heart attacks), and another two may have ambiguous interpretations (S-T segment "straightening" and chest pain as the chief complaint). An improved version of the logistic model [Selker, 1991] used twelve discrete inputs and was shown to perform about as well as an ER physician. To be generally accepted by physicians, however, a decision aid must perform significantly better than physician judgment.

Brush [1984] developed a model based on an "ECG score" that predicted complications in cardiac patients, but the model had disappointing performance when used outside the environment it was developed in [Green, 1988]. Other groups have developed practice guidelines based on expert opinions on how to treat cardiac patients [Weingarten, 1993]. These guidelines work best at selecting patients for early transfer from the CCU, rather than choosing patients suitable for admission.

MODELING TECHNIQUES

Neural Networks

Artificial neural network techniques show excellent promise in being able to overcome the limitations of presently used computer methods to predict patient prognosis. This is because these networks can be trained to recognize complex relationships that exist between inputs (i.e., physiologic data) and outputs (i.e., patient outcome) [de Villiers, 1993]. These subtle relationships in the data are automatically recognized by the network, even if they are unknown to clinicians. Because neural networks can learn any arbitrary relationship between a given set of inputs and outputs, they can normally be expected to perform at least as well as and usually better than any other modeling technique. As the complexity of the problem increases, so does the superiority of neural networks over most other methods. Importantly, neural network techniques have previously been shown to be able to handle the inaccuracy and inconsistency associated with patient histories and physical findings [Pike, 1992; Edenbrandt, 1992; Baxt, 1991; Marik, 1990; Gheorghide, 1988]. Further, the networks appears to be able to deal with the complexities of disease states characterized by several totally differing clinical presentations [Dassen, 1990].

The disadvantage of neural network models is that, while they often have excellent overall results, they do not reveal how a given prediction was made. Physicians sometimes feel uncomfortable with this "black box" approach to patient management in

complicated cases because it is difficult to know when to overrule the network prediction. This objection can be overcome by having a model that can demonstratively perform much better than standard physician judgment.

Rough Sets

Rough sets is a new and powerful technique for extracting rules from data [Pawlak, 1984]. Rough sets have been shown to create impressive predictor models and are especially well suited for problems with inconsistent data, as is often the case with medical problems. Like neural networks, rough sets is a completely data driven technique that can find relationships that exist between problem parameters. A major advantage of rough set models is that they can explain the reason a certain decision was made by revealing what rules were fired. This makes it easier for a physician to reject a decision made by the model on the rare occasions when an unusual set of circumstances suggests such action.

In order to create a rough set model, continuous data must be divided into discrete categories, (e.g., high, medium and low). The rough set algorithm will compare the discretized inputs and output, and eliminate redundant inputs. From the remaining data, a set of rules will be generated that indicates what the likely outcome will be for a given combination of inputs. Certain rules are generated from consistent examples and uncertain rules are generated from inconsistent data. For example, an uncertain rule might state that under given conditions the outcome will be positive 80% of the time. Various methods are employed to give strengths to different rules so that when contradictory rules are fired the most important one will determine the decision.

Rough sets have a few minor disadvantages that have to do with the requirement for discretization of continuous data. If a problem has more than a few inputs, a large amount of data is required to extract rules for all possible combinations of input categories. If a rule has not been generated for a particular combination during training (i.e., rule extraction from a training set of example cases), then no decision can be made when this particular combination occurs during model use. Also, several examples of each combination of categories are desirable to ensure the rules work for a majority of cases. Therefore, a large number of training examples are necessary for the rough set model to generate reliable rules for all possible scenarios.

A second slight disadvantage of rough sets has to do with the crispness of the categories defined for continuous data. For example, a heart rate of 40 - 60 might be considered low, 61 - 80 medium and 81-120 high. Two people may have nearly identical physiologic signs, but one has a heart rate of 80 and the other a heart rate of 81. These people would be considered as being in different categories (80 = Medium, 81 = High), even though they are nearly identical. If a large set of examples is available to extract rules from, this disadvantage can be overcome by using a large number of categories for important variables.

Logistic Regression

Logistic regression is a standard statistical tool that has been used for predictive models with some success [Pozen, 1984; Selker, 1991]. Logistic regression assumes the desired output (usually a "yes" or a "no") fits the sigmoid-shaped logistic equation. The technique has advantages over discriminant analysis in that it can accept combinations of categorical and normal or non-normal continuous data. Data is fit to the equation:

$$Y = \frac{1}{1 + \exp(-u)} \quad (1)$$

where Y is the desired outcome, X are the inputs, b_n are the coefficients of X and $u = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$. Logistic regression has been shown to work well with categorical and non-normal inputs. Its major disadvantage is that it assumes the data fits a rigid form of equation that may not reflect the subtle interactions actually present between factors in the problem.

DATA ANALYSIS

A pilot study, based on an NSF/Whitaker Foundation planning grant, was conducted to determine the feasibility of developing neural network and rough set predictive models from CCU data. A total of 118 records from patient who had gone through the CCU of the University of Arkansas for Medical Sciences' University Hospital in the past five years was input into a database. Surgery patients, catheterization patients and people admitted to the CCU because they are in the midst of a potentially lethal event were excluded. Thirty seven physiologic parameters from the patients charts were recorded, with 28 model inputs recorded at admission and 9 upon admission to the CCU (see Table I). Four possible adverse outcomes were noted: 1. Type II 2nd degree AV block or 3rd degree AV block; 2. More than 15 seconds ventricular tachycardia; 3. Blood pressure less than 85 with the use of pressors; 4. Death. A total of 44 of the patients suffered serious events while in the CCU. Due to the small number of total events, all four adverse outcomes were combined into a single outcome that was positive if any of the four complications occurred.

Model Input Selection

Data from 118 cases was collected, but only 40 of these had a complete set of inputs. The type of data collected creates special problems for model development for several reasons: 1) there are too few training cases for the number of inputs present; 2) the inputs are correlated; and 3) bad data points probably exist in both the inputs and outputs. A set of predictive model inputs was chosen in a two step process. First, data was divided into two groups based on the outcome (yes = event and no = no event). Student t-tests are a method of testing whether the mean of two groups are equal. t-tests were run to look for differences in each variable between the two groups. Afterwards, stepwise logistic regression was run on the variables selected by the t-tests to choose the final set of model variables. The t-tests were necessary because stepwise regression is performed only on cases that have a full set of all inputs. If a single input is missing from a example, then the entire case is removed from the procedure. This, when applied over the entire dataset, then leaves very few complete cases for model development. On the other hand, t-tests can be performed on all cases where the variable under consideration is present, irrespective of whether any of the other inputs are missing. This allows each candidate input to be evaluated over a larger sample size, thus giving a more solid basis for elimination of parameters that show no difference between outcome groups. After candidate inputs are selected by the t-tests, stepwise regression is performed to eliminate redundancies in the inputs caused by correlations between variables.

Eighteen variables were chosen by the t-tests ($p < 0.1$ using either the yes and no groups pooled or separated for calculation of variances) as being possible candidates for the model inputs. The eighteen were: sex, age, weight, diabetes, chest pain, systolic pressure, respiration rate, white blood count, ventricular arrhythmias, ST segment depression, rales, syncope, S3 heart sound, temperature in CCU, diastolic pressure in CCU, respiration in CCU, aspirin use, class III drug use, class IV drug use, and change in body temperature between ER and CCU. After running stepwise logistic regression, seven inputs were chosen for model development: sex, age, weight, diabetes, ST segment

TABLE I. - INPUT PARAMETERS FOR THE PREDICTIVE MODELS.

INPUT #	PHYSIOLOGIC PARAMETER	RANGE
1	sex	male or female
2	age	continuous
3	weight	continuous
4	smoking	yes or no
5	history of diabetes	yes or no
6	previous MI	yes or no
7	chest pain	yes or no
8	heart rate	continuous
9	systolic blood pressure	continuous
10	diastolic blood pressure	continuous
11	body temperature	continuous
12	respiration rate	continuous
13	hematocrit	continuous
14	serum K	continuous
15	white blood count	continuous
16	creatinine	continuous
17	current MI	yes or no
18	anterior MI	yes or no
19	atrial arrhythmias	yes or no
20	ventricular arrhythmias	yes or no
21	ST segment depression	yes or no
22	ST segment elevation	yes or no
23	# of ventricular ectopics in a run	continuous
24	rales greater than 1/3 up	yes or no
25	syncope	yes or no
26	height	continuous
27	S3	yes or no
28	history of congestive heart failure	yes or no
29	heart rate in unit	continuous
30	systolic blood pressure in unit	continuous
31	diastolic blood pressure in unit	continuous
32	respiration in unit	continuous
33	aspirin	yes or no
34	class I drugs	yes or no
35	class II drugs	yes or no
36	class III drugs	yes or no
37	class IV drugs	yes or no

depression, respiration rate in CCU and aspirin use. A total of 95 out of the original 118 cases had all seven of these inputs present.

Factor analysis by principle component decomposition was performed on these seven inputs plus an additional input, presence or absence of atrial arrhythmias, to try to eliminate correlations in the inputs. Three factors were chosen by this method: factor 1 was a combination of sex, respiration rate in the CCU, ST segment depression and diabetes. Factor 2 combined weight, diabetes and atrial arrhythmias, while factor 3 combined aspirin usage and atrial arrhythmias. The resulting factors were fed into a stepwise logistic regression model. The logistic model selected only a constant term, indicating that these three factors have little, if any, predictive power. It was therefore

concluded that factor analysis was not an effective means of reducing this particular dataset.

Training and Testing Set Selection

Model development and validation were performed by dividing the database into two categories, one for model training and the other for model testing. Ideally, a training set should capture the important features in the data. The training set should normally be unbiased (i.e., have an equal number of yes and no outcomes), or be intentionally biased to favor a particular result. It is also desirable to have the testing set representative of the data as a whole, so as to get a true idea of model performance. To accomplish these, the data set was clustered by cases, using a nearest neighbor algorithm. Six clusters were visually identified, with between 2 and 31 members in each cluster. Four cases were far from all others, and these were placed in the test set. Two training sets were developed, one with 61 cases and the other with 40. The set with 40 cases was nearly equally balanced between yes and no answers, while the other one had 24 extra no outcomes. The test set, which contained 33 cases, had all clusters represented and contained 13 positive and 20 negative outcomes.

Neural Network Results

The models created were evaluated by using sensitivity and specificity:

$$\text{sensitivity} = \frac{tp}{tp+fn}$$

$$\text{specificity} = \frac{tn}{tn+fp}$$

where tp is true positives, tn is true negatives, fp is false positives and fn is false negatives. Sensitivity is a measure of how likely a model will predict a condition if it is actually present, while specificity indicates how likely a condition is to be present if the model results are positive. Several neural network architectures were investigated, with the best results shown in Table 2.

TABLE 2.- NEURAL NETWORK RESULTS FOR 7 INPUT MODELS.

	Number of Hidden Nodes		
	3	4	3-1
Average % Correct	58	57	55.5
Sensitivity	0.62	0.54	0.46
Specificity	0.50	0.60	0.65

In Table 2, average % correct is the average of the sensitivity and specificity x 100, while 3-1 indicates a four layer network with three nodes in the first hidden layer and one node in the second hidden layer. The results for three hidden nodes used the training set with 40 cases, while the others used the set with 61 cases.

It was thought that the test set results may have suffered from too many inputs for the number of training cases, so a reduced set of inputs was chosen for further model

generation. The new inputs were age, weight, ST segment depression and respiration rate in the CCU. The training set for this network had 61 cases. A network with 2 hidden nodes had the following results:

Average % Correct = 70.5, sensitivity = 0.46, specificity = 0.95

The results are significant. While the model only correctly predicted about one half of all the cardiac events, when it did forecast an event the patient was extremely likely to suffer one (19 out of 20 cases). This network can therefore be used as a screening tool to help decide to place patients in the CCU or, if they are already in the CCU, to keep them there.

Another technique tried to improve model performance was to combine the outputs of the best networks for sensitivity and specificity. These were used as the inputs for a second neural network, with the idea that if each of the original models searched a different area of the solution space then combining them will produce results better than either alone. The output from the network that had a sensitivity of 0.62 (see Table 2) and the one that had a specificity of 0.95 (described above) were combined. The best architecture had four nodes in a single hidden layer:

Average % Correct = 64.5, sensitivity = 0.54, specificity = 0.75

The results are in between the original networks for sensitivity and specificity, thus indicating that the networks were probably keying in on the same features.

The final method tried was to add simulated training cases in order to increase the allowable degrees of freedom in the problem. This procedure also forces the network to learn relationships between inputs. The procedure is as follows:

1. Calculate an average value over all the cases in the training set for each input.
2. For each case, the number of new exemplars created will equal the number of inputs to the model.
3. Each new exemplar replaces a single input with its mean, so that the number of simulated cases created equals the original number of cases times the number of model inputs.

The procedure described above allows a network to be trained with a larger number of hidden nodes without overtraining the network. The inputs for this model were: sex, age, weight, diabetes, ST segment depression, respiration rate in CCU and aspirin use. The original training set had 41 cases, 19 of which were positive outcomes and 22 negative. The new training set had 328 cases with 152 positive outcomes and 176 negative ones. The best network had a single hidden layer with four hidden nodes:

Average % Correct = 66, sensitivity = 0.57, specificity = 0.75

The results improve upon those shown in Table 2, but are slightly worse (66% vs. 70.5% average correct) and not as useful as those from the network with a reduced set of inputs. The limited number of training cases and the combining of four disparate events into a single outcome probably preclude better model performance on this dataset.

Logistic Regression and Rough Set Results

A logistic model was also developed from the same dataset. The training set with 61 inputs was used for coefficient determination (see Equation 1), and the standard 33 case test set was used for model validation. The best validation results were obtained

with the following inputs: age, weight, ST segment depression, respiration rate in CCU plus the interactions age x ST segment depression, and weight x respiration rate in CCU:

Average % Correct = 64.5, sensitivity = 0.54, specificity = 0.75

The results are not as good as the best neural networks, but better than many of the networks developed. The logistic model therefore is probably a good benchmark to compare the neural network models to, because it gives an indication if the optimal neural network architecture has been developed for a given problem.

A rough set model was developed from four inputs: age, weight, ST segment depression and respiration rate in the CCU. Continuous inputs were divided into four equally spaced categories that spanned their range. Twelve rules were extracted from the 61 case training set, five for negative decisions and seven for positive decisions. The rule certainty was 100% for eleven rules, and 96% for the twelfth. Each negative rule had between four and twenty-five cases supporting it, with positive rules having between one and six cases supporting them. Decisions were made in 31 of 33 cases in the test set. Model results were:

Average % Correct = 73.5, sensitivity = 0.58, specificity = 0.89

These results are excellent compared to logistic regression and neural network techniques. Although the specificity was slightly less than the best neural network model, its overall performance was better. Moreover, the rough set model made no decision in cases that were not similar to those it was developed on, whereas neural networks will always give an output for all cases.

CONCLUSIONS

Rough sets, neural networks and logistic regression have all proven to be effective tools for predicting the outcome of cardiac patients in a CCU. The rough set model gave the best overall results, and has the advantage of being able to explain how a decision was made. Also, rough set models will not make decisions on cases that are far from the ones they were developed on, adding a degree of confidence to the results. The best neural network model proved to be the most practical, with a specificity of 0.95, although overall results were not quite as good as with rough sets. Logistic regression proved useful as a benchmark against which other methods could be tested.

The key to developing these prognostic models is to choose a good set of predictor variables. This was done in a two step process, using student t-tests and stepwise logistic regression. Selection of cases for training and testing models is also crucial for model creation and validation. A clustering algorithm that measures the distance between cases, while requiring subjective decisions, has shown itself to be useful.

Future work includes applying the data analysis techniques described above to the Contextual Alarms Management System (CALMS) project. The goal of CALMS is to develop sophisticated models to predict the onset of clinical cardiac ischemia before it occurs. The system will continuously monitor cardiac patients and set off an alarm when they appear about to suffer an ischemic episode. The models take as inputs information from patient history and combine it with continuously updated information extracted from blood pressure readings, oxygen saturation measurements and five ECG leads. Data is now being collected on twenty patients at the cardiac catheterization laboratory at

Cooper Hospital in New Jersey. Raw data is read into specialized analysis software developed by Po-Ne-Mah. A total of 110 physiologic parameters are written to a text file, which is updated every 1 second. Episodes of ischemia are annotated by physician during the procedure. Since there are too many parameters for the number of patients, each patient will be compared with themselves, with data taken during ischemic episodes compared with data taken when the patient is not suffering ischemia. Student t-tests and logistic regression will be used to choose indicators of ischemia. These will be input into logistic regression, neural network, rough set and expert system models to diagnose and predict future onset of ischemic conditions. One problem that needs to be addressed is drift in these physiologic conditions with time. One possibility for addressing this problem is to look at changes in parameters when ischemia begins, as opposed to absolute readings. Another possibility is to look at inputs in the frequency domain to examine parameters such as heart rate variability and QRS frequency components.

REFERENCES

Baxt, WG. Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction. *Annals of Internal Medicine*. vol. 115, number 11, pp. 843, 1991.

Bianchi AM, Mainardi L, Petrucci E, Signorini MG, Mainardi M, and Cerutti S, Time-variant power spectrum analysis for the detection of transient episodes in HRV signal, *IEEE Transactions on Biomedical Engineering*, Vol 40, 2:136-144, 1993.

De Martini M, Valentini R, Cesana B, Massari FM, Lettino M, Pupilella T, Ambrosini F, Eriano G, La Marchesina U, Lotto A. Early and late prognosis in acute myocardial infarct. A retrospective study in patients admitted to the coronary care unit in the past 10 years. *Giornale Italiano Di Cardiologia* 20:215-26, 1990.

De Villiers J, Barnard B. Backpropagation neural nets with one and two hidden layers. *IEEE Transactions on Neural Networks*. vol 4 no 1. pp 136-141, 1993.

Fineberg HV, Scadden D and Goldman L, Care of patients with a low probability of acute myocardial infarction, *New England Journal of Medicine*, Vol. 310 20:1301-1307.

Gheorghide M, Anderson J, Rosman H, Lakier J, Velardo B, Goldberg D, Friedman A, Schultz L, Tilley B, Goldstein S. Risk identification at the time of admission to coronary care unit in patients with suspected myocardial infarction. *American Heart Journal*. 116:1212-17, 1987.

Green L, and Smith M., Evaluation of two acute cardiac ischemia decision-support tools in a rural family practice, *Journal of Family Practice*, Vol 26, No. 6:627-632, 1988.

Marik PE, Lipman J, Eidelman IJ, Erskine PJ. Clinical predictors of early death in acute myocardial infarction. A prospective study of 233 patients. *South African Medical Journal*. J77:179-82, 1990.

Pawlak Z. Rough sets. *International Journal of Information and Computer Science* 11-2. pp. 341-356, 1982.

Rollag, AJ, Jonsbu, Aase, O, and Erikssen, J, Standardized use of simple criteria from case history improves selection of patients for cardiac care unit admission, *Journal of Internal Medicine*, 232:299-304, 1992.

Selker, HP, "Coronary unit triage decision aids: how do we know they work?", American Journal of Medicine, 87:491-493, 1989.

Selker HP, Griffith JL and D'Agostino RB, A tool for judging coronary care unit admission appropriateness, valid for both real-time and retrospective use, Medical Care, Vol 29, 7:610-627, 1991.

Weingarten SR, Ermann B, Reidinger MS, Shah PK and Ellrodt AG, Selecting the best triage rule for patients hospitalized with chest pain, American Journal of Medicine, 87:494-500, 1989.

Weingarten SR, Ermann B, Bolus R, Reidinger MS, Rubin H, Green A, Karns K and Ellrodt AG, Early "step-down" transfer of low risk patients with chest pain, Annals of Internal Medicine 113:283-289, 1990.

Weingarten SR, Agocs L, Tankel N, Sheng A and Ellrodt AG, Reducing lengths of stay for patients hospitalized with chest pain using medical practice guidelines and opinion leaders, American Journal of Cardiology, 71:259-262, 1993.

**SOME CONSIDERATIONS IN THE DETERMINATION OF THE
ACCURACY OF A MEASUREMENT IN SPACE OF THE
NEWTONIAN GRAVITATIONAL CONSTANT (G)**

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1995

Johnson Space Center

Prepared By:	Stephen D. Baker, Ph.D.
Academic Rank:	Professor
College and Department:	Rice University Physics Department MS-61 6100 South Main Street Houston, TX 77005-1892
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Earth Science and Solar System Exploration
Branch:	Planetary Missions and Materials
JSC Colleague:	Wendell W. Mendell
Date Submitted:	October 23, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

A commonly suggested method for determining the Newtonian constant of universal gravitation (G) is to observe the motion of two bodies of known mass moving about each other in an orbiting laboratory. In low Earth orbit (LEO), bodies constructed of even the densest material available experience a gravitational attraction that is several times smaller than the "tidal" forces (due to their proximity to the Earth), which tend to pull them apart. While the tidal forces do not preclude stable orbits of the two objects about each other, they and the Coriolis force (in the rotating laboratory) dominate the motion, and the gravitational attraction of the two bodies may be considered a weak (but significant) contribution to the motion. As a result, compared to an experiment that would be performed in a laboratory far from the Earth, greater accuracy of measuring the motion of the two bodies may be required for a given accuracy in the determination of G . We find that the accuracy with which positions must be determined is not much different in an experiment in LEO than in one performed far from the Earth, but that rotational periods must be determined more accurately. Using a curvature matrix analysis, we also find that a value of G may be extracted (with some loss in accuracy, but probably some practical gain) from an analysis of the time dependence of the distance between the bodies rather than of a full specification (distance and direction) of their relative positions. A measurement of the gravitational constant to one part in 10^4 continues to be thinkable, but one part in 10^5 will be very difficult.

INTRODUCTION

A commonly suggested method for determining the Newtonian constant of universal gravitation (G) is to observe the motion of two bodies of known mass moving about each other in an orbiting laboratory. In low Earth orbit (LEO), bodies constructed of even the densest material available experience a gravitational attraction that is several times smaller than the "tidal" forces (due to their proximity to the Earth), which tend to pull them apart. While the tidal forces do not preclude stable orbits of the two objects about each other, they and the Coriolis force (in the rotating laboratory) dominate the motion, and the gravitational attraction of the two bodies may be considered a weak (but significant) contribution to the motion. As a result, compared to an experiment that would be performed in a laboratory far from the Earth, greater accuracy of measuring the motion of the two bodies may be required for a given accuracy in the determination of G . In this report, we show that the accuracy with which positions must be determined is not much different in an experiment in LEO than in one performed far from the Earth, but that rotational periods must be determined more accurately. In the final section of this report, we briefly describe a curvature matrix analysis which shows that a value of G may be extracted (with some loss in accuracy, but probably some practical gain) from an analysis of just the time dependence of the distance between the bodies rather than of a full specification (distance and direction) of their relative positions.

If a system of two balls is far from any other objects, the balls can move around each other in elliptical Keplerian orbits, where the relationship between the gravitational constant G , the period T of the motion, the average (half the sum of the minimum and the maximum) separation a of the balls, and the total mass m of the system is given by

$$G = 4\pi^2 a^3 / mT^2 \quad (1)$$

This leads us to the following result. If a , T , and m are measured with accuracies Δa , ΔT , and Δm , respectively, we may determine the fractional error in G from the fractional errors in a , T , and m as follows:

$$\left(\frac{\Delta G}{G}\right) = 2\left(\frac{\Delta T}{T}\right) + 3\left(\frac{\Delta a}{a}\right) + \left(\frac{\Delta m}{m}\right) \quad (2)$$

Thus, to achieve a given fractional accuracy in the determination of G , the fractional accuracy of the period must be at least two times better, the fractional accuracy of the orbit size must be at least three times better, and the fractional accuracy of the mass must be at least as good.

SIMPLE ORBITS IN LOW EARTH ORBIT

For the motion of two balls in an orbiting laboratory in a low Earth orbit (LEO), we must take into account the tidal forces and the Coriolis forces and integrate numerically the equations of motion (assuming the laboratory is in a circular orbit about a spherically symmetric Earth). It turns out that two balls may still orbit about one another in a stable fashion, but the motion is complicated by the presence of the Earth. I will lay out the equations of motion here, and I will then comment on them in the following paragraph.

The equations of motion are

$$\begin{aligned}
 \ddot{x} &= \dot{z} - \rho x [x^2 + y^2 + z^2]^{-3/2} \\
 \ddot{y} &= -y - \rho y [x^2 + y^2 + z^2]^{-3/2} \\
 \ddot{z} &= 3z - 2\dot{x} - \rho z [x^2 + y^2 + z^2]^{-3/2}
 \end{aligned} \tag{3}$$

where x is along an axis in the direction of motion of the orbiting laboratory, z is along an axis directed toward the Earth from the laboratory (NASA convention), and y is the third Cartesian coordinate. The distances x , y , and z are measured in dimensionless units relative to a characteristic length d in the laboratory, time is measured in dimensionless units relative to the reciprocal of the angular velocity of the laboratory about the Earth, and the quantity ρ is defined:

$$\rho \equiv (m/M)(A/d)^3 \tag{4}$$

where m is the mass of the two-ball system, M is the mass of the Earth, and A is the radius of the laboratory's orbit about the Earth.

The tidal forces (the first terms on the right side of the expressions for \ddot{y} and \ddot{z}) and the Coriolis forces (the velocity dependent terms) can be large compared to the gravitational attraction (the last term in all three expressions). It turns out that if the two balls are of equal mass and are made of material the density of tungsten, and if d (the length scale) is chosen to be the distance between their centers when they are in contact, the quantity ρ is very nearly equal to unity in LEO (for sintered tungsten balls of 10 kg each, density about 19.1 g/cm³, the distance between their centers when they are touching is about 10 cm.) Thus, in LEO, for real balls moving about one another, the gravitational attraction term will be small compared to the average values of the other terms in Eq. (3), and the motion will be only gently (but for our purposes still importantly) affected by the gravitational attraction between the two balls.

An example of a possible stable orbit of two balls about each other is shown in Figure 1, which shows the motion of one ball relative to the other. Parameters of the motion are described in the caption. Generally, the orbits (unlike the Keplerian orbits) are not closed. Moreover, the initial conditions must be chosen carefully to give motion in which the particles remain close for a long time. Fig 2 shows an example of the relative motion of the balls in which they do not stay close together for long. For the purposes of the next section, it was necessary to determine the parameters for the special case of closed orbits, several of which are shown in Fig 3. Here we see that large orbits have periods approaching 2π and a motion which is simple harmonic in x and in z , with a ratio of amplitudes of 2. Smaller orbits are more nearly circular and have shorter periods. They are ovals but they are not ellipses.

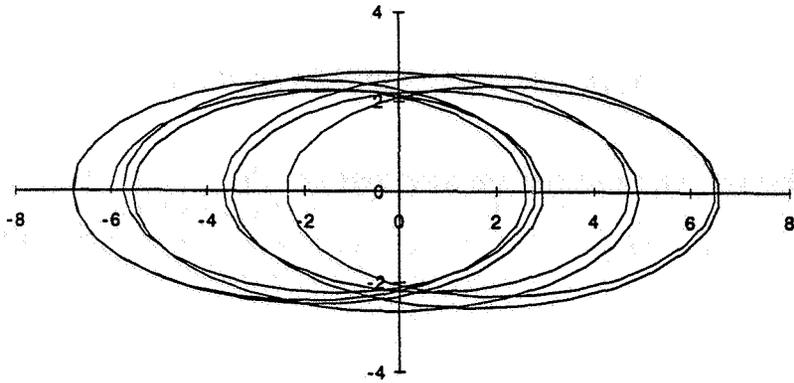


Figure 1.— Relative position in xz plane of two balls moving about each other in LEO. The x direction (horizontal axis) is in the direction of motion of the laboratory, and the z direction (vertical axis) is toward the Earth from the laboratory. $\rho = 1$, $x_0 = -6.0$, $v_{0x} = 0.1$, $v_{0z} = 2.5$, all other initial values of coordinates and velocity components are zero. The motion is stable in the sense that the balls remain within about 7 length units indefinitely. The trajectory shown is for the time interval between $t = 0$ and $t = 40\pi/3$ (6.67 orbits of the laboratory about the Earth).

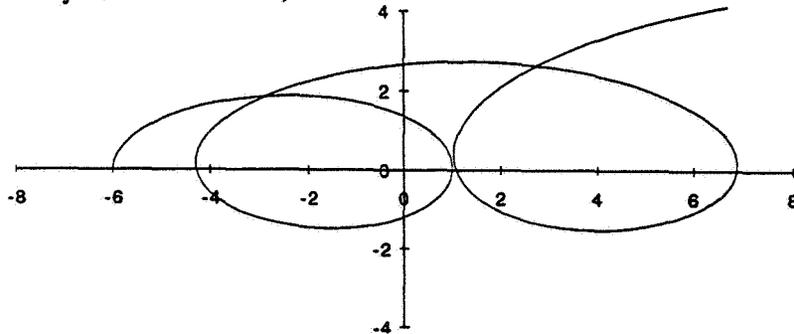


Figure 2.— Relative position in xz plane of two balls moving about each other in LEO. $\rho = 1$, $x_0 = -6.0$, $v_{0x} = 0.1$, $v_{0z} = 2.1$, all other initial values of coordinates and velocity components are zero. The motion is unstable, and after just one loop around the origin, the trajectory continues to the right in a cycloidal fashion. The trajectory shown is for the time interval between $t = 0$ and $t = 4\pi$ (about two orbits of the laboratory about the Earth).

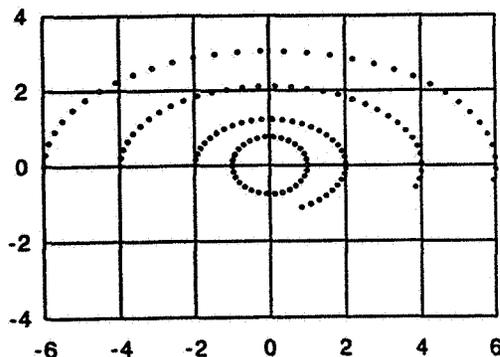


Figure 3.— Relative position in xz plane of two balls moving about each other in LEO. $\rho = 1$. The initial values of coordinates and velocity components have been chosen to give closed orbits. The time interval between dots is 0.10; for the largest orbits, the period is close to 2π , for smaller orbits the period is smaller.

ERRORS ASSOCIATED WITH SIMPLEST ORBITS

Equations (3) tell us that when the particles are close together, the gravitational attractions between them is more important than the tidal and Coriolis forces. (While this is not possible in LEO, it would be possible if the laboratory were farther from the Earth.) The motion of the two balls is in Keplerian orbits, and the relationship between period (dimensionless units), maximum separation a (dimensionless units), and ρ is

$$a^3/T^2 = \rho/4\pi^2 \quad (5)$$

where $\rho = (m/M)(A/d)^3$. Note the similarity to the expression for the Kepler motion of an isolated binary; Gm is simply replaced by ρ .

The task of measuring G becomes the task of measuring ρ . From the above relation, Eq.(5), we see that, far from the Earth,

$$\left(\frac{\Delta\rho}{\rho}\right) = 2\left(\frac{\Delta T}{T}\right) + 3\left(\frac{\Delta a}{a}\right) \quad (6)$$

The issue is: How does $\left(\frac{\Delta\rho}{\rho}\right)$ depend on $\left(\frac{\Delta T}{T}\right)$ and $\left(\frac{\Delta a}{a}\right)$ when the experiment is carried out in LEO? That is, what happens to the factors 2 and 3 in the equation above? (Of course, getting from ρ to G also involves knowing the mass of the binaries, the length scale, and the radius of the orbit of the laboratory about the Earth. Since we should be able to know these to the $1:10^6$ level, we will ignore them now, so the accuracy with which ρ is determined is practically the accuracy with which G is determined.)

Here is a description of the calculation. With ρ set equal to 1, for each of several values of a (maximum separation) we found the orbit which was closed and determined the period of the orbit. We also determined numerically the partial derivative of T with respect to a , $(\partial T/\partial a)$, and the partial derivative of T with respect to ρ , $(\partial T/\partial \rho)$. We can then determine the fractional change in ρ which results from given fractional changes in a and T , as follows.

We may write

$$dT = \frac{\partial T}{\partial a} da + \frac{\partial T}{\partial \rho} d\rho, \quad (8)$$

which may be rearranged to give

$$d\rho = \frac{dT}{\left(\frac{\partial T}{\partial \rho}\right)} - \frac{\left(\frac{\partial T}{\partial a}\right)}{\left(\frac{\partial T}{\partial \rho}\right)} da \quad (9)$$

Dividing all terms by ρ and multiplying numerator and denominator of the two terms on the right side of the equal sign by T or a , respectively, we obtain

$$\left(\frac{\Delta\rho}{\rho}\right) = \left(\frac{dT}{T}\right) \left(\frac{T}{\rho \left(\frac{\partial T}{\partial \rho}\right)}\right) - \left(\frac{da}{a}\right) \left(\frac{a \left(\frac{\partial T}{\partial a}\right)}{\rho \left(\frac{\partial T}{\partial \rho}\right)}\right) = \mathcal{A} \left(\frac{dT}{T}\right) + \mathcal{B} \left(\frac{da}{a}\right) \quad (10)$$

Eq. (10) indicates how the coefficients \mathcal{A} and \mathcal{B} are calculated. If the orbits were Keplerian (small a , assuming ρ is unity), the values of \mathcal{A} and \mathcal{B} would be 2 and 3, respectively, as we see in Eq. (6). The actual values for larger, and potentially usable values of a , are shown in Figure 4. What we find is that the ρ becomes much more sensitive to the time measurement. This is reasonable because as the orbits become larger, the effect of the gravitational attraction of the masses becomes smaller, and the period becomes practically independent of the size of the orbit. What may be surprising (it is surprising to me) is that the fractional precision with which the size of the orbit must be determined varies very little from the value of 3 for the values of a in our region of interest.

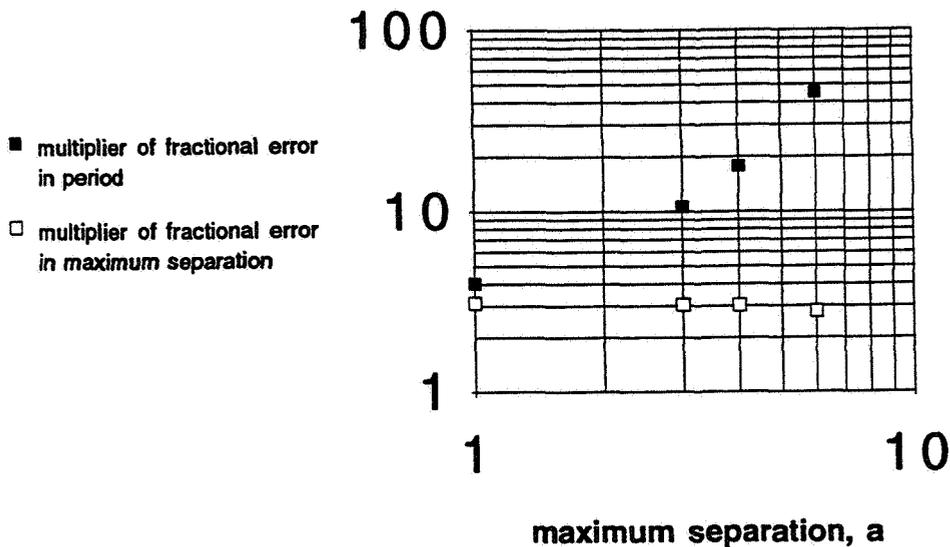


Figure 4.— Values of the coefficients \mathcal{A} (solid squares) and \mathcal{B} (open squares) versus the size of the orbit of the relative motion of the two balls. See text for explanation.

We can now understand the effect of LEO on the accuracy with which we need to measure the orbital motion to extract G . What we find is that the period must be measured with much greater fractional accuracy, whereas the size of the orbit still needs to be measured with about the same fractional accuracy as before.

CURVATURE MATRIX ERROR ANALYSIS

In the analysis of an actual experiment, data points, consisting of position and time measurements, will be fit to a numerical model of the motion. The parameters of the model (initial position and velocity, strength of the gravitational interaction, effect of other masses) are determined by a least squares search process and the precision of the parameters is deduced from the curvature of the chi-square hypersurface in parameter space. We may first consider a slightly simpler problem. We calculate the orbits with parameters and produce positions at regular intervals of time. By changing each of the parameters and each pair of parameters by small amounts, we can calculate a chi-square which characterizes the deviation of the modified set of positions from the original set. From these chi-squares, we determine the curvature matrix which describes the dependence of chi-square on the parameters. Inverting this curvature matrix then gives an indication of the sensitivity of the motion to each of the parameters, in particular to the gravitational attraction between the members of the orbiting binary. One would expect this analysis, which is more sophisticated than the analysis of the simplest, closed orbits, described in the previous section, will agree with the error analysis of the simple closed orbits but can be extended to orbits which are not closed and which reflect more complex environments.

We have analyzed an orbit of two 10 kg tungsten balls with parameters $d = 10$ cm, $\rho = 1$, $x_0 = -6.0$, $v_{0x} = 0.1$, $v_{0z} = 2.1$, all other initial values of coordinates and velocity components being zero. Supposing the position is determined at 100 points, each with a precision of about $10 \mu\text{m}$, at time intervals of 0.25 (for a total time of about 6 orbits of the laboratory about the Earth), we find that the precision with which ρ (for these assumptions) is determined is about 2.7×10^{-5} , consistent with the simpler analysis described above. This indicates that a determination of G to one part in 10^4 is thinkable (but one part in 10^5 will be difficult).

It is interesting to note that one may also perform such an analysis with the chi-square calculated not from the relative positions of the particles (separation distance and direction) but from the deviation of just the separation of the particles (in analogy with the analysis of lunar laser ranging data). Some information is lost in doing this, but for the same conditions as in the previous paragraph, the precision with which ρ is determined is about 4.4×10^{-5} , not much worse than before.

In conclusion, a measurement (of the kind discussed above) in LEO of the gravitational constant to one part in 10^4 continues to be thinkable, but one part in 10^5 will be very difficult.

**MATERIALS ASSESSMENT OF COMPONENTS OF THE
EXTRAVEHICULAR MOBILITY UNIT**

**Final Report
NASA/ASEE Summer Faculty Fellowship Program-1995
Johnson Space Center**

Prepared By:	Enrique V. Barrera, Ph.D.
Academic Rank:	Assistant Professor
and:	John D. Olivas
	Graduate Student
University and Department:	Rice University
	Mechanical Engineering and Materials Science
NASA/JSC	
Directorate:	Engineering
Division:	Crew and Thermal Systems
Branch:	
JSC Colleague:	Chin Lin, Ph.D.
Date Submitted:	August 9, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

Current research interests for Extravehicular Mobility Unit (EMU) design and development are directed toward enhancements of the Shuttle EMU, implementation of the Mark III technology for Shuttle applications, and development of a next generation suit (the X suit) which has applications for prolong space flight, longer extravehicular activity (EVA), and Moon and Mars missions. In this research project two principal components of the EMU were studied from the vantage point of the materials and their design criteria. An investigation of the flexible materials which make up the lay-up of materials for abrasion and tear protection, thermal insulation, pressure restraint, and etc. was initiated. A central focus was on the thermal insulation. A vacuum apparatus for measuring flexing of the materials was built to access their durability in vacuum. Plans are to include a Residual Gas Analyzer on the vacuum chamber to measure volatiles during the durability testing. These tests will more accurately simulate space conditions and provide information which has not been available on the materials currently used on the EMU. Durability testing of the aluminized mylar with a nylon scrim showed that the material strength varied in the machine and transverse directions. Study of components of the EMU also included a study of the EMU Bearing Assemblies as to materials selection, engineered materials, use of coatings and flammability issues. A comprehensive analysis of the performance of the current design, which is a stainless steel assembly, was conducted and use of titanium alloys or engineered alloy systems and coatings was investigated. The friction and wear properties are of interest as are the general manufacturing costs. Recognizing that the bearing assembly is subject to an oxygen environment, all currently used materials as well as titanium and engineered alloys were evaluated as to their flammability. An aim of the project is to provide weight reduction since bearing weights constitute 1/3 of the total EMU weight. Our investigations have shown favorable properties using a titanium or nickel base alloy in conjunction with a coating system. Interest lies in developing titanium as a more nonflammable material. Methodology for doing this lies in adding coatings and surface alloying the titanium. This report is brief and does not give all necessary details. The reader should contact the authors as to the detailed study and for viewing of raw data.

INTRODUCTION

A broad-based project aimed at studying the flexible materials and the bearing assemblies on the Extravehicular Mobility Unit (EMU) was initiated. The emphasis of the study of the flexible materials became the thermal insulation layers of aluminized mylar with nylon scrim and its durability in flexure testing. The emphasis on the bearing assemblies was focused on flammability and improving the flammability of titanium and its alloys. Both components to this project were aimed at current Shuttle EMU material systems, applying Mark III technology, and criteria of the next generation suit, the X-suit. The outcome of this project is a plan in place for flexure testing of the flexible materials used on the Shuttle EMU where the materials evaluation occurs in vacuum to simulate space conditions. The plan is presented in this report as well as the assessment of the aluminized mylar. Recommendations for reducing frequent replacement of the aluminized mylar are also included. For the study on the bearing assemblies several recommendations are presented and methodology for further assessment is also given. In this summer program the faculty fellow and student participant focused on accomplishing the initial stages to a hopefully continued study. The objective of the program is to maintain a continued relationship where NASA interests are fulfilled. The report outlines solutions for that goal.

VACUUM DURABILITY TESTING OF EMU FLEXIBLE MATERIALS

Objectives

The durability and breakdown resistance of fabric materials currently used on the Shuttle Extravehicular Mobility Unit (EMU) will be determined using a Flex Machine developed during the 1995 NASA-ASEE Summer Faculty Fellowship Program. The Flex Machine is designed to simulate the flexing movements made by the astronauts during extravehicular activity (EVA) in vacuum conditions resembling that of low earth orbit (LEO). The tester is designed to work in vacuum and to minimize gas evolution from the fixture. Volatile gases will be measured during the testing. Gases that evolve during the tests are a product of the material degradation. As a result, the findings of this study will be used to improve materials that see frequent replacement or repair and to aid in selecting materials for prolonged EVA and time in space. Both current Shuttle EMU and X-suit materials will be evaluated. The benefits of conducting the tests in vacuum are that the volatile gases that would outgas in space can be measured and the modification/degradation of the materials being exposed to vacuum can be induced for observation by electron microscopy, microprobe analysis, and x-ray photoelectron spectroscopy. It is apparent that knowing the properties of the material degradation due to use in vacuum will further play a role for space suits left on the space station or that make trips to the Moon and Mars.

Aluminized mylar with a nylon scrim, the current thermal insulation material was tested in tension and in flexure modes in ambient conditions. The aluminized mylar failed before the nylon scrim and the machine direction was significantly stronger than the transverse direction. Expectations are that the method of processing the material system results in reduced strength in the transverse direction. Optical micrographs of the material showed lines in the material resembling Lüder's bands. These features are under continued study. The scrim shows two different conditions where nylon is twisted tight in one direction and is laid loose in the perpendicular direction. It is suspected that this feature does not alter the failure mode of the aluminized mylar but effects the percent elongation of the part in terms of final failure. The loose nylon elongates more before failure. The adhesive used on the thermal insulation causes the aluminized mylar to show draw up

possibly due to shrinkage. This may impart small creases in the aluminized mylar which are associated with the low strength failure in the transverse direction.

Background

Current flexible materials on the Shuttle EMU include eight layers of various materials for flame, abrasion, and tear resistance, thermal insulation, micrometeoroid protection, and pressure restraint [1]. The current materials and the selection process itself has gone through an evolution since the suits for the Mercury program were first designed, placing more and more emphasis on lighter weight, material stability, and extended durability[2]. Vacuum testing of space suit materials was conducted as far back as 1964 and continues as full mock up testing [3]. Current test apparatus at JSC allow for vacuum testing but do not provide for volatile gas measurement and extended materials analysis. Furthermore, durability tests are run in ambient conditions which serve as a *Safe Life* test (materials designed such that no failure will take place in the design lifetime). Many materials currently used and those being considered for the X-suit are composite materials, meaning that they take advantage of the properties of a number of component materials which make up one part or fabric [4]. These materials allow for some built in redundancy.

Research plan

Flex testing in vacuum will be conducted on current Shuttle flexible materials to evaluate durability for Shuttle and prolonged use. Temperature control will be put on the chamber during the project year. Extended materials analysis will be conducted on the vacuum tested materials. Material outgassing conditions will be mathematically modeled. Further design of current materials will be conducted in collaboration with Crew and Thermal Systems personnel and contractor companies. Emphasis is placed on improving the performance of the EMU while setting up a criteria for materials selection based on gas evolution during use in space (simulated on earth).

In the summer program, preliminary tests were conducted on the aluminized mylar in ambient conditions and the flex tester for vacuum testing was built. The following list of deliverables demonstrates the methodology by which this research will be continued and the long term goals of this study.

August 8, 1995:	Project starting date. Test and redesign frequently repaired or replaced materials.
October 15, 1995	Submission of a Regional University Grant Proposal: "Development of thermal and radiation insulation".
January 15, 1996	Report I: Interim report.
February 15, 1996	Submission of an unsolicited proposal: "Development of composite material for weight savings and functionality".
May 14, 1996	Project ending date.
June 14, 1996	Final Report: Including expenditures

Recommendations

The aluminized mylar with a nylon scrim is the current choice for thermal insulation on Shuttle EMU. To reduce the number of repairs currently seen for the Shuttle suit care

must be taken in aligning the material with the machine direction in the direction of the major loads. Cross plying sounds like a possible solution yet will result in failure of half the layers as opposed to failure of all the layers. Recommendations of this study are to align the machine direction with the direct of maximum loading. This will lead to a longer life of the insulation and no failure in the transverse direction.

MATERIAL SELECTION FOR EMU BEARING ASSEMBLIES

In analyzing the application of a material system to the EMU bearing assembly, the selections were evaluated based on the qualifying parameters listed in Table 1. Prior to evaluating potential material systems, baseline data of the current material selection was compiled to provide a starting point for evaluation. This criteria established the minimum needed to justify making a material change. The material properties are seen in Table 2.

TABLE 1.- REQUIREMENTS FOR EMU BEARING MATERIAL SELECTION

Property	Consequence
Low weight	Minimize mass
High stiffness	Minimal deflection
	Minimal torsional distortion
Non flammable in 100% O ₂	Minimum 4.3 psia
	Maximum 6.6 psia
Manufacturing ease	Easily machined
	Easily cast
	Available in stock blanks
Good wear characteristics	Bearing race application
	Bearing ball application
Impact resistant	High fracture toughness
Commercial application	Non-aerospace applications
Cost	To be determined based on selections

TABLE 2.- BASELINE DATA FOR EXISTING BEARING MATERIALS [3]

Material	Bearing Component	Treatment	Tensile Strength (MPa/ksi)	Combustion 4.3-6.6 psi O ₂	Thermal Expansion (x10 ⁻⁶ /°C)	Hardness (HRC)	Density (lbs./in ³)
440C	Balls	Temper @ 300 °C	1970/285	No	10.1	60	0.275
17-4PH	Race rings	H900	1310/190	No	10.4	40-48	0.28

Given the bearing requirements, the most stringent is compatibility of the material with an oxygen enriched environment [5]. There are two engineering components to the issue of combustion; ignition source and material combustion. Non-flammable materials are those in which combustion is not supported in an oxygen enriched environment, as defined by NASA White Sands Test Facility (WSTF) [6]. Risk of ignition can be minimized by reducing the potential for spark or flame ignition. Ideally both aspects should be satisfied, however, significantly reducing one may increase the ability to use the material. In order to evaluate a greater number of materials, the investigation was divided into three material classes; metals, ceramics and composites. A number of materials of

each class were evaluated based on the above criterion. Recommendations based on material class follow each section as well as an overall recommendation.

Metals evaluation

Investigation of potential metal selections for bearing applications was performed using published ASM data [7] and available vendor information. Realistic metal systems which were initially identified as having a density lower than that of stainless steel (17-4 PH) were titanium, aluminum, graphite, beryllium and their associated alloys. Previous designs eliminated aluminum as a potential choice due to low stiffness and graphite due to low toughness. Although beryllium has an excellent strength to weight ratio, poor fracture toughness and toxicity also eliminated this material as a potential selection [8].

Titanium and titanium alloys were determined to be the best potential selection among metal systems. A comparison of several commercially available titanium alloys are listed in Table 3. These represent those alloys which possess comparable properties to the baseline data in Table 1.

TABLE 3.- RELEVANT MECHANICAL PROPERTIES FOR TITANIUM ALLOYS

Material	Bearing Component	Treatment	Tensile Strength (MPa/ksi)	Combustion 4.3-6.6 psi O ₂	Hardness (HRC)	Density (lbs./in ³)
Ti6Al4V	Race	Annealed	895/130	Yes	36	0.16
	Race	Solution	1035/150	Yes	39	0.16
Ti6Al6V2 Sn	Race	Solution	1030/150	Unknown	39	0.165
Ti7Al4Mo	Race	Solution	1170/170	Unknown	32-38	0.162
Ti6Al2Sn4 Zr6Mo	Race	Solution	1170/170	Unknown	36-42	0.168
Ti6Al2Sn2 Zr2Mo2Cr	Race	Solution	1160/168	Unknown	42	0.165
Ti10V2Fe 3Al	Race	Solution	1275/185	Unknown	50	0.168

Draw backs to this material selection class are in the area of combustion. WSTF demonstrated clearly that titanium, Ti-6Al-4V, and several other titanium alloys provided the poorest combustion performance for all metals tested [9]. Recognizing that the primary alloying agent is highly combustible, it is expected that other systems would perform comparably, although different phases appear to play a role in the combustion of some systems.

It is acknowledged that the bearings would not likely be exposed to direct flame contact as simulated in WSTF material combustion test. Testing a proposed bearing in the configurational and component test may provide sufficient support to warrant the use of titanium in this application, however, with regard to the initial criteria, titanium cannot be recommended in the as commercially available conditions without the risk of being consumed in a combustion condition. In conclusion, of the metals currently available and which have been tested by WSTF, none can be recommended without a compromise in either mechanical or combustion properties.

Ceramics evaluation

Several important properties are characteristic of this class material. In general, they are more stable at higher temperature, have high strengths and low weights, are suitably hard, and somewhat machinable. However, drawbacks include the potential for

low fracture toughness (brittle), varying degrees of porosity, poor surface finish, and difficulties in some fabrication processes. The ceramics which are presented possess the most favorable of these initial concerns and comply with the requirements as defined in Table 1 unless otherwise stated.

Ceramic bearings and bearing elements (balls, races, etc.) have been fabricated since the early 1980's and a significant amount of work has been reported in the literature. Although problems were encountered early in the development of the material system, many of the obstacles have been overcome. Table 4 represents the properties of the most common and well studied structural ceramic material for bearing and load carrying applications.

TABLE 4.- RELEVANT MECHANICAL PROPERTIES FOR CERAMICS

Material	Bearing Component	Trade Name	Flexural Strength (MPa/ksi)	Fracture Toughness $\sqrt{\text{in}}$	Combustion 4.3-6.6 psi O ₂	Thermal Expansion ($\times 10^{-6}/^{\circ}\text{C}$)	Hardness (HRC)	Density (lbs./in ³)
Si ₃ N ₄	Race/Ball	NBD-200	980/142	6.4	No	2.9	>70	0.115
Sialon	Race/Ball	NT-451	920/133	6.4	Unknown	3.6	>70	0.117
Si ₃ N ₄	Race/Ball	NT-154	910/132	6.4	No	3.9	>70	0.117
B ₂ C	Race/Ball	NORBIDE	300/44	2.8	Unknown	5.8	>70	0.091
Y-ZrO ₂	Race/Ball	YZ-110	1400/203	8.3	Unknown	10	>70	0.219
Y-ZrO ₂	Race/Ball	YZ-130	1000/145	4.6	Unknown	10	>70	0.218
Y-ZTA	Race/Ball	AZ-67	900/131	6.4	Unknown	8.5	>70	0.159
Y-ZTA	Race/Ball	AZ-93	1180/171	5.5	Unknown	9	>70	0.174
Ce-ZTA	Race/Ball	CAZ-94	650/94	7.4	Unknown	9	>70	0.188

Of the ceramics listed, the only material for which WSTF combustion test data exists is Si₃N₄ (NBD-200) [8]. This performed very well and appears to satisfy all the characteristics defined in Table 1. Mass fabrication and production of bearing parts and bearing assemblies have proven to be successful in many commercial applications [13-19]. For example, ceramic bearings were used successfully in the LOX turbo pump on STS-70. Strides which the ceramic industry have made in the last 10 years have been significant in resolving the problems with using these materials in bearing applications. Based on published information and communication with material vendors, the sizes required by the EMU can be fabricated. In order to keep cost down on fabrication of raw material, it may be necessary to consider some slight alterations in the current design.

Composite evaluation

Most of the composite systems which rely on internal reinforcement for increased strength characteristics will intuitively possess the same problems as do the metals. That is, the exposure of a combustible material to an environment conducive for combustion will lead to material and component failure. Regressing from these conventional ideas, a composite can also be designed by applying a protective coating developed to protect the base material from the combustion environment. Thus the susceptibility of the bulk material to failure can effectively be eliminated. Furthermore, the coating (or film depending on thickness) could be tailored to specifically satisfy other demands placed on the surface. Since the 1950's, the bearing industry has been using a variety of diffusion and deposited coatings on load bearing surfaces to increase hardness, reduce wear, or increase corrosion resistance [20-22].

Recalling from the materials outlined in Section 2, titanium alloys would be an excellent choice provided that one could prevent the material from coming into contact with an ignition source or combusting. Therefore, proper coating selection and design may additionally satisfy the wear, friction, strength, and material compliance requirements so that titanium alloys could safely be used in the oxygen enriched environments. The coated titanium alloy assembly will result in a weight savings over the stainless steel systems currently used.

Since titanium alloys are inherently self passivating and very corrosion resistant, published coating technology has been limited to increasing surface hardness [23]. Similarly, since titanium alloys do not possess the type of hardnesses traditionally found in bearing materials, it represents only a small portion of total data available from bearing applications in contrast to more popular metals such as 440C or M-50 steels. However, several deposition techniques exist (eg., Chemical Vapor Deposition, Physical Vapor Deposition, Sublimation Deposition, etc.) and the design of such a coating is possible with further research. Table 5 demonstrates differences in surface hardness of substrates and coatings the properties of some typical coating used in bearing applications. Additionally, significant data exists on wear properties of various coating combinations. Presented in Table 6 are friction and wear characteristics of a few coating combinations [22].

Many non-combustible metals and alloys are applied to steel alloys for enhanced wear and protection characteristics, such as chrome and nickel electroplating. It is also acknowledged that both are less combustible than titanium. While no previous work of these system on titanium was discovered, applications are possible. Similarly, elements such as cobalt and copper metals and alloys are not combustible in oxygen enriched environments and may prove to be a potential coating selection.

TABLE 5.- HARDNESS OF COATINGS AND RELEVANT METALS USED IN BEARING APPLICATIONS

Material substrate/coating	Hardness (HV)
440C stainless steel (HRC 60)	697
17-4 PH stainless steel (HRC 47)	471
Ti-6Al-4V (HRC 36)	354
Hard chrome plating	1000-1200
Nitrided steel	1300-1700
WC+Co	1400-1800
TiN	2000
Ruby, sapphire, corundum	2500-3000
SiC	4000
B ₄ C	5000
Diamond	>10000

TABLE 6.- PIN ON DISK FRICTION AND WEAR CHARACTERISTICS OF VARIOUS COATING COMBINATIONS [22]

Contacting Surfaces Pin-Disc	Friction Coefficient	Wear Rate	
	Humidity 0.5%-5%	Pin	Disc
TiC-TiN	0.18	4.5	20
TiC-SiC	0.26	0.33	<3
TiC-TiC	0.32	0.25	16
TiN-TiC	0.31	0	25
SiC-TiC	0.35	0	36
SiC-SiC	0.47	6	22
Al ₂ O ₃ -TiC	0.37	0.7	20
	Humidity 50%		
100Cr6-SiC	0.23	1.2	3.8
100Cr6-TiC	0.25	0.07	7.6
100Cr6-TiN	0.49	95	0
100Cr6-X205 CrWMoV121	0.53	104	0
100Cr6-Fe ₃ B	0.76	58	0
100Cr6-Cr ₇ C ₃	0.79	1.1	76
Al ₂ O ₃ -100Cr6	0.45	10	1500
Al ₂ O ₃ -TiC	0.19	0.1	10
Al ₂ O ₃ -boronized cement carbide	0.62	3.3	1.8
TiC-TiC	0.14	4.4	
TiN-TiN	0.19	4.3	
Cr ₇ C ₃ -Cr ₇ C ₃	0.29	29.3	
Fe ₃ B-Fe ₃ B	0.40	0.3	
Ruby-TiC	0.12	4	
100Cr6-TiC	0.11	3	

Recommendations

Although a material which satisfies all the requirements listed is not readily available, a number of opportunities exist in solving the problem. The quickest solution may lie in fabricating the bearings from Si₃N₄. While bearings approaching the size of the waist bearing assembly are not fabricated in bulk, current machining technology may be able to accommodate the design. Major manufacturers such as Timeken and SKF may have capabilities in fabricating these components. Additionally there are two manufacturers (in Japan and Germany) who specialize in ceramic bearings and are accustomed to special orders. They may also be able to aid in redesigning to accommodate more standard sizes and special considerations in using ceramics.

Another opportunity exists in trying to alter the microstructure of an existing titanium alloy or design a new one. Chemical composition being held constant, gamma titanium has demonstrated better combustion characteristics than alpha, alpha-beta, or beta microstructures as measured by WSTF. This dependency on microstructure gives rise to the need to understand the mechanisms behind combustion of solid metals. Since theoretical models do not adequately predict this behavior [23-29], it is recommended to conduct research into developing a more suitable mathematical model of combustion. This work can then be implemented on developing a more suitable material with optimal stable phases which suppress the combustion process and maintain desired strength levels. While

it is not likely that this option will completely resolve the combustion problems, it may sufficiently alter the characteristic such that certain materials may be used.

Developing a composite system using a bulk titanium alloy and combustion protective coating presents another attractive solution. As mentioned, these alloys are rarely coated, therefore, research would be required to investigate which systems would be compatible. There has been sufficient work done in this field to demonstrate that such a coating would be possible. Additionally, some degree of control on specific characteristics such as hardness, wear, and friction, can be exercised in the new design.

DISCUSSION

Many of the materials used as flexible materials on the EMU are polymeric and in turn exhibit some outgasing in a vacuum condition. Short term outgasing is expected to be either minimum in intensity or to happen quickly at the onset of vacuum. Knowledge of the long term exposure to vacuum of the current materials is unknown. Continued degradation of less than thirty years or for durations considered long term exposure limit these materials for use on the X-suit. This is the reason for vacuum flex testing and volatile gas measurement. Any of the solutions recommended for the bearing assemblies would have broad appeal to commercial industry. While specific to the EMU, reporting the results of this application will be immediately evaluated by industries where light weight, high strength, and combustion potential exists. These may include, but are not limited to, gas turbine, chemical processing, and utility distribution applications. For any the recommendations selected, a sound theoretical model would go a long way toward understanding fundamental kinetics and thermodynamics which drive the combustion process. The development of noncombustible titanium would impact the space program as a whole since there are a number of applications where good strength to weight titanium in the nonflammable condition would be of interest.

CONCLUSIONS

A study was conducted on aluminized mylar, the thermal insulation on the Shuttle EMU and bearing assemblies used on the various EMU designs. Recommendations for improved performance were presented for both problems and long range studies were outlined to further contribute to enhanced EMU design.

REFERENCES

1. Shuttle EMU Space Suit Assembly Materials and Application Functions, provided by E. Orndoff, NASA-JSC.
2. R. D. Kozloski, U.S. Space Gear: Outfitting the Astronaut, Smithsonian Institution Press, Washington (1994).
3. Space Environment Test System, AVCO model 354, Contract no. NAS9-3625. (AVC), Tulsa Division.
4. K. K. Chawla, Composite Materials, Springer-Verlag, New York, 1987.
5. R. Christianson and J. M. Stoltzfus, *Proc. ASME Win Ann Mtg*, (1993).
6. S. Sircar, J. Stoltzfus, C. Bryan, and J. Kazaroff, *ASTM STP 1267*, 7 (1995).
7. K. Mills (ed.), *Metals Handbook*, 9th ed., 2 (1987).
8. L. Schoenman and J.E. Franklin, *ASTM STP 1111*, 5 (1991) 354.
9. M. Gunaji, S. Sircar and H. Beeson, *ASTM STP 1267*, 7 (1995).
10. S. Jahanmir, *Mat Res Soc Symp Proc*, 140 (1989) 285.
11. E. Kingsbury, *Mat Res Soc Symp Proc*, 140 (1989) 437.

12. R. N. Katz and J. G. Hannoosh, *Int J High Technology Cermics*, **1** (1985) 69.
13. T. E. Fischer and H. Tomizawa, *Wear*, **105** (1985) 29.
14. J. W. Lucek, *Gas Tur Aero Cong Exp*, (1990).
15. G. W. Hosang, *AIAA 23rd Joint Prop Conf*, (1987) 1.
16. H. Tomizawa and T. E. Fischer, *ASLE Trans*, **30** 1 (1986) 41.
17. D. L. Allen, *ASLE Trans*, **37** 2 (1986) 410.
18. S. Jahanmir and T. E. Fischer, *ASME/ASLE Trib Conf*, (1986) 1.
19. H. Aramaki, Y. Shoda, Y. Morishita and T. Sawamoto, *J Trib*, **110** (1988) 693.
20. J. L. Spodnik and J. J. Wert, , *ASME/ASLE Trib Conf*, (1986) 309.
21. A. K. Gangopadhyay, M. E. Fine and H.S. Cheng, , *ASME/ASLE Trib Conf*, (1986) 293.
22. H. E. Hintermann, *Thin Solid Films*, **84** (1981) 215.
23. S. Sircar, D. B. Wilson and J. Stoltzfus, *Met Trans B*, **24B** (1993) 395.
24. F. J. Benz and J.M. Stoltzfus, *ASTM STP 910*, **2** (1986) 39.
25. F. J. Benz, R. C. Shaw, J. M. Homa, *ASTM STP 910*, **2** (1986) 135.
26. W. W. Yuen, *ASTM STP 910*, **2** (1986) 59.
27. S. Sircar, D. B. Wilson, S. Hornung and J. Stoltzfus, (1995) in press.
28. T. A. Steinberg, S. Sircar, D. B. Wilson, J. M. Stoltzfus, (1995) in press.
29. D. B. Wilson, (1995) in press.

REGIONAL AND LOCALIZED DEPOSITS ON THE MOON

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1995

Johnson Space Center

Prepared By:	Cassandra R. Coombs, Ph.D.
Academic Rank:	Assistant Professor
University & Department	College of Charleston Department of Geology Charleston, S.C. 29424
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
JSC Colleague:	David McKay, Ph.D.
Date Submitted:	August 3, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

Earth-based telescopic remote sensing studies have provided important information concerning lunar pyroclastic deposits. Combined with the returned lunar sample studies and analyses of lunar photography, we have learned a great deal about the nature and origin of these explosive volcanic materials. Lunar pyroclastic deposits are more numerous, extensive, and widely distributed than previously thought. Two generic classes of lunar pyroclastics have been identified, regional and localized. From the former, two separate spectral compositional groups have been identified; one is dominated by Fe^{2+} -bearing glasses, the other is composed of ilmenite-rich black spheres. Comparatively, three separate spectral groups have been identified among the localized deposits: highlands-rich, olivine-rich, and mare-rich. Returned sample studies and the recently collected Galileo and Clementine data also corroborate these findings. Albedo data and multispectral imagery suggest that the thicker core deposits of the regional dark mantle deposits (RDMD) are surrounded by pyroclastic debris and subjacent highlands material. The presence of a major component of pyroclastic debris in the regolith surrounding the core regional deposits has important implications for the resource potential of these materials. Both telescopic and orbital spectra indicate that the regional pyroclastic deposits are rich in iron, titanium and oxygen-bearing minerals. Particle shapes vary from simple glass spheres to compound droplets with quench crystallized textures. Their small grain size and friability make them ideal indigenous feedstock. Compared to other resource feedstock sources on the Moon, these pyroclastic materials may be the best oxygen resource on the Moon.

INTRODUCTION

Explosive volcanic, or pyroclastic, materials are unique phases in the lunar soils and are important as they hold clues to the history of lunar volcanism. As impact craters can be used as windows to the lunar interior, volcanic deposits can be used as windows to the deep lunar interior. Craters excavate and uplift materials from depths ranging from the near-surface to as much as 1/10 the crater diameter (Pieters *et al.*, 1994). Pyroclastic glasses, among the most chemically "primitive" of lunar rocks, directly sample depths as great as 400 km (Delano, 1986) and thus can help address two major science theme strategies put forth by the Lunar Exploration Science Working Group, LExSWG: to better understand the formation of the Earth-Moon system and the thermal and magmatic evolution of the Moon (LESWG, 1992), while helping us to plan better for upcoming lunar missions.

Earth-based telescopic studies have provided most of our information concerning lunar pyroclastic deposits. Combined with the returned lunar sample studies and analyses of spacecraft photography, we continue to gain insight into the nature and origin of these explosive volcanic materials. Two generic classes of lunar pyroclastics are recognized: regional and local. Our previous work has shown that the larger regional deposits are more numerous, extensive, and widely distributed than previously thought (e.g., Coombs, 1988; Hawke *et al.*, 1989; Coombs & Hawke, 1995), leading us to suggest that they may exhibit distinct compositional variations. Returned sample studies and the recently collected Galileo and Clementine spacecraft data also corroborate these findings (e.g., Greeley *et al.*, 1993; McEwen *et al.*, 1994).

Whole and broken green glass beads were first found in abundance at the Apollo 15 site. On the Apollo 17 mission, orange glass beads and their quench-crystallized black equivalent were found in high concentration at Shorty Crater Station 4 (Figure 1). Various interpretations for the origin of these glass beads were proffered: (1) impact melt ejecta from large impacts which had penetrated to more mafic material at depth (Carr and Meyer, 1972; Hussain, 1972), (2) vapor condensates (Cavaretta *et al.*, 1972; Quaide, 1973), (3) splash droplets from impacts into lava lakes (Roedder and Weiblen, 1973), or (4) pyroclastic material (McKay and Heiken, 1973; McKay *et al.*, 1973; Huneke, 1973; Carter *et al.*, 1973; Reid *et al.*, 1973; Heiken and McKay, 1974). The explosive volcanic, or pyroclastic origin is now commonly accepted for the Apollo 17 orange and black glass spheres and Apollo 15 green glass. Similarly, it is now commonly accepted that the widely spread dark mantle deposits present elsewhere on the Moon also formed in similar explosive eruptions.



Figure 1: Trench at Apollo 17 revealing orange soil. Samples collected from this area also collected black glass spheres.

Since their initial identification and classification, many more dark mantle deposits have been identified. Among the largest of these are the Aristarchus Plateau, Mare Humorum, South Mare Vaporum, and Sulpicius Gallus deposits. Characteristically, the lunar pyroclastic deposits are very smooth, low albedo (0.079 - 0.096; Pohn and Wildey, 1970) units that cover and subdue underlying terrain. The regional pyroclastic deposits occur as thick accumulations in topographic lows and are thin or absent on adjacent hilltops (e.g., Lucchitta and Schmitt, 1974). Visual observations by Apollo astronauts and 2m-resolution Orbiter photographs indicate that the surface of these deposits is relatively fine textured with a velvety smooth appearance (Lucchitta, 1973; Schmitt, 1974; Lucchitta and Schmitt, 1974; Gaddis *et al.*, 1985).

Color varies among the deposits from a bluish-gray to a reddish-brown and locally may have a strong red, orange or blue hue (e.g., Schmitt, 1974).

During the past decade an enlightened attitude developed toward man's exploration and colonization of space. As part of this reawakening, the possibility of establishing a manned lunar base needs to be revisited. When established, such an endeavor should provide sound economic benefits as well as being a solid base from which to conduct scientific experiments. In the immediate post-Apollo era it was suggested by a number of workers that titanium production might be a profitable economic activity for a permanent

needs to be revisited. When established, such an endeavor should provide sound economic benefits as well as being a solid base from which to conduct scientific experiments. In the immediate post-Apollo era it was suggested by a number of workers that titanium production might be a profitable economic activity for a permanent manned lunar base. Following that, it was suggested that fine-grained lunar regolith material would be useful for shielding orbiting space stations or military facilities. In the past few years, attention has focused on the production of oxygen propellant and helium-3 as nuclear fusion fuel both for use at the base and to transport materials back to Earth (e.g., Simon, 1985; Gibson and Knudson, 1985; Kulcinski *et al.*, 1986; Kulcinski, 1988). From these early studies, ilmenite-rich material became the preferred source for the production of these resources (see Mendell, 1985; Hawke *et al.*, 1989a,b). Hawke *et al.* (1990) discussed the resource potential of lunar pyroclastic deposits and suggested that they would make an excellent site for locating a lunar base. Most recently, Allen *et al.* (1995) extracted 6% O₂ from Apollo 17 glass spheres, the highest amount collected yet. In this paper we summarize some of what is known about these deposits, discuss models for their emplacement, and further explore their scientific, engineering and resource-related advantages as lunar base sites.

LUNAR PYROCLASTIC MATERIALS

REGIONAL VS LOCAL

Based on recent geologic and remote sensing data, the lunar pyroclastic deposits have been divided into two genetic classes: regional and localized (e.g., Lucey *et al.*, 1984; Gaddis *et al.*, 1985; Coombs, 1989; Hawke *et al.*, 1989). These subdivisions are based upon their compositions as well as overall size and distribution. The regional dark mantle deposits cover a relatively large area of the lunar surface with respect to the smaller, localized deposits. Indeed, the regional deposits vary in size from 4,000 to 30,000 km². These large deposits are located in highlands areas adjacent to (and in some cases are superposed on) many of the major maria. The explosive fire-fountaining that formed these regional deposits may have been associated with some of the early mare-filling volcanic episodes (McGetchin and Head, 1973; Head, 1974). Endogenic craters and other irregular depressions are thought to be the source vents for these deposits (e.g., Zisk *et al.*, 1973; Head, 1974; Gaddis *et al.*, 1985; Coombs and Hawke, 1988).

The localized pyroclastic deposits are more widely spread across the lunar nearside than are the regional pyroclastics. These deposits too were emplaced via volcanic fire-fountaining. On average, the localized deposits cover 250-550 km² although some may be as small as 80 km², and some as large as 700 km² (Coombs *et al.*, 1987; Coombs, 1988; Hawke *et al.*, 1989). They are concentrated around the perimeters of the major lunar maria and are commonly found in the floors of large Imbrian and pre-Imbrian aged impact structures (e.g., Head and Wilson, 1979; Coombs and Hawke, 1988). The localized deposits are generally associated with small endogenic source craters (<3 km) and are aligned along crater floor-fractures and/or regional faults. Typically, these source craters are irregular in shape and lack obvious crater rays.

The explosive eruptions that formed the lunar dark mantle deposits have been likened to some types of terrestrial volcanic activity. The regional deposits most likely formed in a manner similar to terrestrial strombolian fire-fountaining, while the smaller, more localized deposits most likely formed by "vulcanian-type" explosions (Wilson and Head, 1979; Head and Wilson, 1981; Coombs, 1988; Hawke *et al.*, 1989). A strombolian, or continuous eruption cycle is consistent with the volatile coated spheres returned from the Apollo 17 landing site, which chemical studies have shown to have originated deep in the lunar interior (<300 km).

In a strombolian eruption relatively small time-transient explosions occur at intervals ranging from less than 0.1 second to hours. Gas and pressure build up as the magma rises to the surface. The rising magma eventually reaches the surface through propagating cracks and or faults in the overlying rock. Explosive decompression occurs as the pressure is released and the magma and gas rise in an expanding column of erupting material. Environmental conditions on the Moon cause the particles to spread out over an area roughly six times larger on the Moon than they would for a similar eruption on Earth. With the 1/6 gravity and lack of atmosphere on the Moon, no pyroclastic flows will occur, rather, the pyroclasts will be spread over a broad area, with a size-grading of the clasts. Larger fragments will be deposited closest to the vent while finer grained particles may travel 100's km before settling out (Wilson and Head, 1979).

In the vulcanian-type eruptions that formed the localized pyroclastic deposits, magma, gas and pressure build up beneath a cap-rock in the conduit. Eventually, when enough pressure builds up, explosive decompression occurs and the cap-rock is blown away, along with the rising magma and gas. These eruption columns too will spread out more evenly on the Moon and cause the settling pyroclasts to cover a larger area than they would on Earth (Coombs, 1988; Coombs *et al.*, 1989).

Due to their small areal extents and relative thinness, however, it is thought that the smaller, localized deposits would be less efficient for use in extracting lunar resources, and thus will be left out of the current discussion and evaluation as a potential lunar resource.

Pyroclastic Soil Evolution

Sample work done on the returned Apollo 17 orange and black glasses indicates that they are chemically indistinguishable from each other; the only difference being that the black spheres are partially quench-crystallized. Collectively, the glass beads are fine grained and have a mean grain size of ~40 μm and their shapes vary from subequant spheres to angular fragments. Many of them are coated with sublimates and or micromounds. Fig. 2 illustrates the breakdown of the mean grain size versus standard deviation for the A17 soils. From this figure, it is clear that the orange and black glasses are finer grained than most mature lunar soils.

Figure 2: Mean grain size versus standard deviation for lunar soils. Note where the Apollo 17 orange glass lies in relation to other lunar soils. It is much finer grained than even mature lunar soils.

A size equilibrium is reached within lunar soils as a result of meteorite reworking. Basically, the grain size of any soil is determined by the amount of fresh ejecta present. For example, if there is a lot of fresh ejecta present, the soil will be fairly coarse grained. Or, if not much fresh material is introduced, the grains will become smaller and smaller over time as the mature soil is continuously being reworked by micrometeorites and impacts. Eventually, the soil or regolith is made into agglutinates that in turn, are reworked and mixed together with extraneous exotic materials. These materials then continue to be reworked. In time, it appears that with thicker soils, such as the 30 - 50 m thick dark mantle deposits, the steady addition of ejecta from fresh bedrock will go away and the continuous reworking of the already present regolith is continued.

If the supply of fresh material diminishes, the grain sizes will reduce to approximately 20 μm . Earlier studies by McKay *et al.*, (1974), indicate that this 20 μm size is the smallest fraction that the glass beads to which will break down. The grain size is determined by the ratio of large to small particles. If there is no longer a significant input of larger particles, then the grain size will continue to decrease and reach an equilibrium floor of ~20 μm .

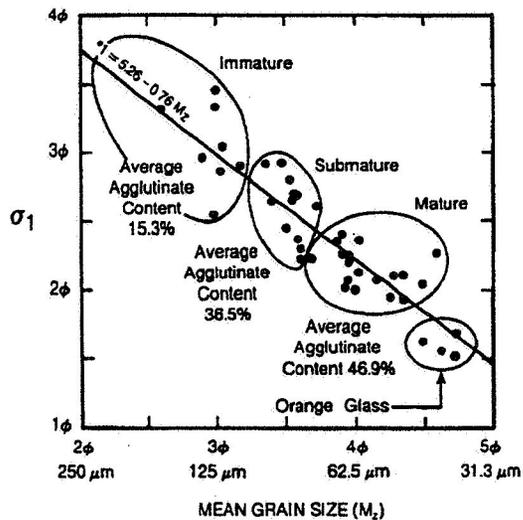
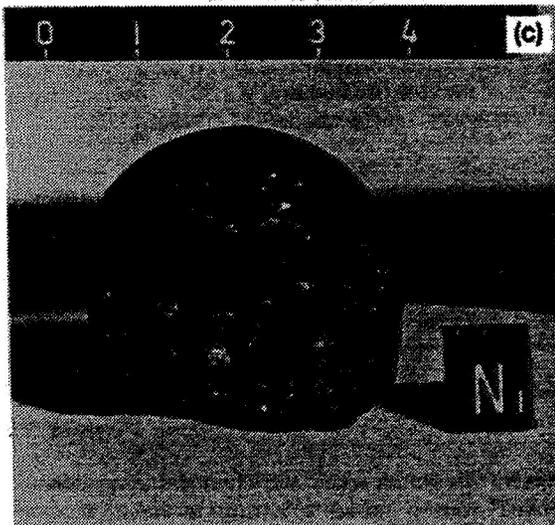


Figure 3: Vesicles within glass sphere 60095 from returned Apollo 17 sample suite. Clast size is approximately 1 cm. Vesicles due to presence of internal gas during formation and cooling phase.

The maximum estimated thickness for a regional pyroclastic deposit is ~50 m. Based on early work by McKay and Heiken (1974) and Carrier (1974) the mean grain size within this deposit will be 5.35 ϕ or 25 μm . For an average thickness of 30 m, the mean grain size expected for a mature deposit would be 29 μm or 5.11 ϕ . Analyses of the entire orange and black glass collection have yielded no single particle size greater than 1 mm, although some fragile clods of agglutinate material were larger than 1mm.

Morphology of Pyroclasts

Looking a little more closely at these glasses, the particle shapes vary from simple glass spheres to compound droplets with quench crystallized textures that vary from fine-grained dendrites to subequant skeletons. The bulk of the returned glass samples is comprised of fine grained fragments and shards, which raises a question about their origin and emplacement. These fragments are now interpreted to be the result of an explosive eruption, as mentioned above, whereby the material erupted, landed, bounced and more material was deposited on top. As a result, a process of self collision and breakdown occurs. These fragments are beneficial to the resource potential of these glasses. That is, the more surface area exposed, the more solar wind that can be adsorbed by the glasses. Quick calculations indicate that the surface area of a small bucket full of lunar pyroclastic materials is the equivalent surface area of an American football field (K. Joosten, pers. communication).

Vesicles are seen in less than 6% of the returned samples. Vesicles form in pyroclastic material by trapping gases in the magma during the fire-fountaining phase. The gases become isolated and trapped as the magma blebs cool and solidify during flight. These gases may represent volatiles released during the volcanic activity, and thus may be representative of the interior composition of the Moon. Possible vesicle gases include fluorine compounds, sulfur, solar-wind gases and carbon monoxide (Goldberg *et al.*, 1976). Carbon monoxide remains among the favorite of the vesicle-forming gases. Laboratory studies performed by Gibson *et al.* (1975) found that most of the C in mare basalts is released as CO and CO₂ as trapped magmatic gases. This, in addition to atmospheric and sputtering losses and recycling into subsequent lava flows, may account for the low C content currently detected in lunar soils.

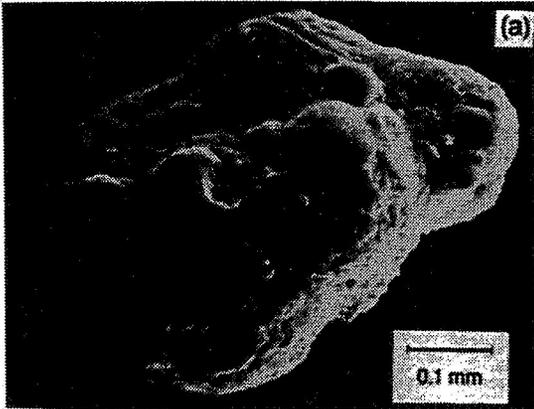


Figure 4: Compound droplet collected from the 74001/2 core tube sampled at Apollo 17.

Many of the glass beads form compound droplets (Figure 4), whereby one bead cooled and another hot droplet hit the bigger one while it was still soft and "squished" around it. Some of these secondary or parasitic droplets have been shown to be partly crystallized just beneath the surface. It is thought that the film, composed of Zn, Ga, Pb, Cu, Ti, S, F, Cl and other elements condensed on the

surfaces during lava fountaining. Meyer *et al.* (1975) propose an anhydrous sulfide, chloride, and fluoride rich vapor, originating from a pyroxenite layer deep inside the Moon. In addition, many of the glass beads are coated with a thin film of micromounds (Figure 5). The micromounds range in size from 20 - 500 Å in diameter (McKay *et al.*, 1973). These are interpreted to be vapor condensate features that formed by venting gases, similar to the condensate deposits found around terrestrial volcanic/fumarolic centers. The micromounds found on the lunar glasses are enriched in sulfur, with some Zn, K, and Na also present (Clanton *et al.*, 1978).

Morphologic textural features present within the Apollo 17 glass sample suite include the predominance of olivine and ilmenite crystals. These are thought to be due to devitrification of the glass bead. Parasitic crystals have also been found on the surfaces of some beads. The exact composition of these crystals is unknown, however, the larger ones are predominately of two compositions, potassium-chloride and sodium-chloride. The surfaces of some orange glass beads were also found to contain amoeboid blebs of iron.

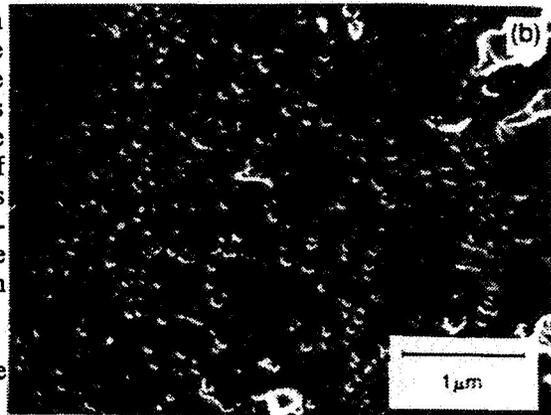


Figure 5: SEM photomicrograph of sublimate micromounds on the surface of a glass sphere.

Table 1: Pristine Lunar Glass Varieties Arranged According to wt % TiO₂ Abundance

CLASS/VARIETY	SiO ₂	TiO ₂	Al ₂ O ₃	Cr ₂ O ₃	FeO	MnO	MgO	CaO	Na ₂ O	K ₂ O
1) Apollo 15 Green C	48.0	0.26	7.74	0.57		0.19	18.2	8.57	n.d.	n.d.
					16.5					
2) Apollo 15 Green A	45.5	0.38	7.75	0.56	19.7	0.22	17.2	8.65	n.d.	n.d.
3) Apollo 16 Green	43.9	0.39	7.83	0.39	21.9	0.24	16.9	8.44	n.d.	n.d.
4) Apollo 15 Green	46.0	0.40	7.92	0.55	19.1	n.a.	17.2	8.75	n.d.	n.d.
5) Apollo 15 Green D	45.1	0.41	7.43	0.55	20.3	0.22	17.6	8.43	n.d.	n.d.
6) Apollo 15 Green E	45.2	0.43	7.44	0.54	19.8	0.22	18.3	8.15	n.d.	n.d.
7) Apollo 14 Green B	44.8	0.45	7.14	0.54	19.8	0.24	19.1	8.03	0.06	0.03
8) Apollo 14 VLT	56.0	0.55	9.30	0.58	18.2	0.21	15.9	9.24	0.11	0.07
9) Apollo 11 Green	43.7	0.57	7.96	0.46	21.5	n.a.	17.0	8.44	n.d.	n.d.
10) Apollo 17 VLT	45.3	0.66	9.60	0.40	19.6	0.26	15.0	9.40	0.27	0.04
11) Apollo 17 Green	44.3	0.91	6.89	n.a.	20.2	0.23	19.5	7.40	0.10	n.d.
12) Apollo 14 Green A	44.1	0.97	6.71	0.56	23.1	0.28	16.6	7.94	n.d.	n.d.
13) Apollo 15 Yellow	42.9	3.48	8.30	0.59	22.1	0.27	13.5	8.50	0.45	n.d.
14) Apollo 14 Yellow	40.8	4.58	6.16	0.41	24.7	0.30	14.8	7.74	0.42	0.10
15) Apollo 17 Yellow	40.5	6.90	8.05	0.63	22.3	0.25	12.6	8.64	0.39	n.d.
16) Apollo 17 Orange	39.4	8.63	6.21	0.67	22.2	0.28	14.7	7.53	0.41	0.04
17) Apollo 17 Orange	38.5	9.12	5.79	0.69	22.9	n.a.	14.9	7.40	0.38	n.d.
18) Apollo 15 Orange	37.9	9.12	5.63	0.65	23.7	n.a.	14.9	7.41	0.36	n.d.
19) Apollo 17 Orange	38.8	9.30	7.62	0.66	22.9	0.29	11.6	8.55	0.39	n.d.
20) Apollo 11 Orange	37.3	10.0	5.68	0.63	23.7	n.a.	14.3	7.62	0.31	n.d.
21) Apollo 14 Orange	37.2	12.5	5.69	0.86	22.2	0.31	14.5	7.04	0.28	0.29
22) Apollo 15 Red	35.6	13.8	7.15	0.77	21.9	0.25	12.1	7.89	0.49	0.12
23) Apollo 14 Red	35.6	15.3	4.81	n.a.	23.7	n.a.	13.0	6.49	0.50	n.d.
24) Apollo 14 Black	34.0	16.4	4.6	0.92	24.5	0.31	13.3	6.90	0.23	0.16
25) Apollo 12 Red	33.4	16.4	4.6	0.84	23.9	0.30	13.0	6.27	0.05	0.12

n.a.: not yet analyzed

n.d.: not detected

Modified from Delano (1986).

GEOCHEMISTRY OF THE LUNAR PYROCLASTIC DEPOSITS

Geochemistry Of Returned Samples

Since their return, many microprobe, SEM, TEM, and petrographic analyses have been done on the lunar glass droplets (e.g., Chao *et al.*, 1970; Reid *et al.*, 1972, 1973; Ringwood, 1973; Heiken and McKay, 1974; Green *et al.*, 1975; Glass, 1976; Marvin and Walker, 1978; Warner *et al.*, 1979; Green and Delano, 1980; Delano and Livi, 1981; Chen *et al.*, 1982; Delano, 1986). As a result of these many detailed studies Delano (1986) identified and confirmed the existence of twenty-five varieties of pristine, or volcanic, glass. These classes are based on their individual compositions and include: Apollo 11 green and orange; Apollo 12 red; Apollo 14 green (A & B), VLT, yellow, orange, black, and red; Apollo 15 green (A, B, C, D, E), orange, yellow, and red; Apollo 16 green; and Apollo 17 orange, yellow, VLT, and green (see Table 1). To accomplish this, Delano synthesized all of the previously published data and removed any interlaboratory bias by consistently using the Apollo 17 pristine orange glass composition as a working standard throughout his investigations of the high-Ti glasses. Likewise, the Apollo 15 pristine yellow glasses and Apollo 15 pristine green A glasses, respectively, were used as standards for other glass droplets containing medium- and low-Ti abundances (Delano, 1986).

The laboratory data indicate a composition rich in SiO₂, FeO and MgO for the various pyroclastic classes. These values are consistent with remote sensing data collected over the various regional pyroclastic deposits and discussed below.

REMOTE SENSING OF LUNAR PYROCLASTIC DEPOSITS

Orbital Geochemistry

The Apollo 15 and 16 spacecrafts each flew a fluorescent X-ray experiment package from which intensity ratios for the elements Mg, Al and Si were measured. In a procedure outlined by Bielefeld (1977), the various intensity ratios were converted to geochemical concentration ratios by weight. From these data, a positive correlation was noticed between high Mg/Al ratios and the presence of dark mantle material (Schonfeld and Bielefeld, 1978). Of the four large, regional pyroclastic deposits mentioned thus far, Sulpicius Gallus has the highest overall average Mg/Al values. Studies by Butler *et al.* (1978) show that the green, orange and brown pyroclastic glasses have extremely high Mg/Al concentration ratios (2.7, 2.6 and 1.7, respectively) compared to ordinary mare soils (ave. 0.61). When these glasses are mixed with the mare soils, an overall increase in Mg and decrease in Al concentration occurs. Mixing models performed by Schonfeld and Bielefeld (1978) estimate that approximately 32-37% of the Sulpicius Gallus formation is composed of orange and black glass. Similarly, they estimate that the two glasses compose 28% of the Taurus-Littrow deposit. Returned samples, however, indicate that only about 10-25% of the Apollo 17 soil was composed of orange and black glass, with the exception of Shorty Crater where pure concentrations of orange and black glass were found (AFGIT, 1973). This discrepancy may be due to the 50 km distance between the Apollo 17 landing site and the Taurus-Littrow deposit.

Other similar correlations between high Mg/Al ratios and the presence of dark mantle pyroclastic material were found at the craters Picard (Olson and Wilhelms, 1974) and Pierce (Casella and Binder, 1972) in Mare Crisium as well as in Mare Fecunditatis.

Andre *et al.* (1975) also analyzed the orbital geochemistry data collected over the Apollo 17-Taurus Littrow site. They looked at the Al/Si concentration and found that a decrease in intensity of the Al/Si ratio correlates with a distinct chemical boundary between the highland and dark mantle units. Less pronounced, however, was the distinction between the dark mantle unit and the floor of Mare Serenitatis. Here, the X-ray fluorescence data indicated only a slight variation in Al/Si between the two materials (Andre *et al.*, 1975). Nine Taurus-Littrow basalt samples studied by Nava (1974), Rhodes *et al.* (1974), and Rose *et al.* (1974) have an Al/Si concentration ranging from 0.22 to 0.28, or an intensity ratio between ~0.50 to ~0.66. The average concentration ratio yields an Al/Si intensity ratio of ~0.59, which corresponds to the average value for four orbital data points in the mare proper (Andre *et al.*, 1975). The calculated concentration and intensity ratios for the orange and black glass are 0.19 and 0.48 respectively. These data correspond nicely with the visible albedo data of Pohn and Wildey (1970), the color-difference photograph of Whitaker (1972), and the relative spectral reflectivity curves of McCord *et al.* (1972), Gaddis *et al.* (1985), Coombs (1988), and Hawke *et al.* (1989).

Earth-Based Radar

As discussed briefly above, 3.8- and 70-cm radar backscatter images show that the surfaces of the regional pyroclastic deposits lack signal-scatterers, or rocks and boulders in the 1-50 cm size range. Rather, the data indicate that the surfaces of these deposits are smooth and block-free, and, that the nearsurface zones of these deposits are also block-free (Zisk *et al.*, 1974; Thompson, 1979). With the low angle of incidence in the 3.8-cm radar images, the polarized signals returned are largely a function of the local slope. The unpolarized signals, on the other hand, are almost entirely dependent upon the inherent properties of the surface materials, and exhibit little or no slope effects. These data support the field observations made by Apollo 17 astronauts, Cernan and Schmitt, during their EVA at Shorty Crater (Bailey and Ulrich, 1975), where they cored into a dark mantle deposit. In general, within any particular dark mantle deposit, areas of high radar return, or where a small impact penetrated the pyroclastic mantle and exposed higher albedo material, are rare.

Earth-Based Telescopic Spectra

Earth-based telescopic spectral reflectance studies have provided important information concerning the composition of regional pyroclastic deposits (e.g., Adams *et al.*, 1973; Pieters, 1973, 1974; Gaddis, 1985; Lucey *et al.*, 1986). Figure 6 shows a number of near-infrared spectra that were collected from regional pyroclastic deposits on the lunar nearside (e.g., Sinus Aestuum, Rima Bode, Mare Vaporum, Taurus Littrow, Aristarchus Plateau). These spectra were collected from the University of Hawaii 2.2 meter telescope on Mauna Kea and processed according to the methods outlined by McCord *et al.*, 1981.

The spectra indicate that the regional pyroclastic deposits are rich in iron, titanium and oxygen-bearing minerals. Some minor compositional variations do exist between the deposits. For example, the top three spectra in Figure 6 from Taurus Littrow, Sinus Aestuum, and Rima Bode, indicate a composition rich in iron, titanium and oxygen. The absorption bands in the continuum removed spectra are shallow and broad, with somewhat irregular base curves. The bottom two spectra from, Mare Humorum and the Aristarchus Plateau, on the other hand indicate that these deposits are rich in Fe^{2+} -bearing glass (Gaddis *et al.*, 1985; Lucey *et al.*, 1986; Coombs, 1988; Hawke *et al.*, 1989). Here the absorption bands exhibit broader, deeper and more smooth or uniform signatures and extend to longer wavelengths.

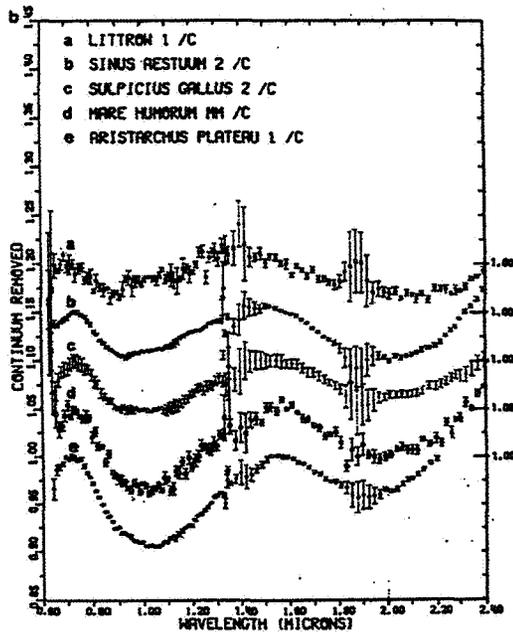


Figure 6: Continuum removed spectra collected from the U.H. 2.2 m telescope on Mauna Kea of some regional pyroclastic deposits.

Multispectral ratios of these deposits exhibit very high 0.40/0.56 μm values and the deposits appear "blue" in 0.40/0.56 μm images presented by Pieters *et al.* (1974) and McCord *et al.* (1976).

RESOURCE POTENTIAL OF LUNAR PYROCLASTIC MATERIALS

The surfaces of the lunar dark mantle deposits are uniform and exhibit a low amount of contamination (Pieters *et al.*, 1974; McCord *et al.*, 1976, 1979; Gaddis *et al.*, 1981, 1985; Coombs, 1988; Hawke *et al.*, 1989). The deposits are widespread and cover large areal extents. They are relatively deep, varying from 10-50 m thick. The deposits are homogeneous with uniform sizes and compositions and are relatively unconsolidated.

The mechanical properties of the pyroclastic deposits are predictable as compared to typical mare or highland regoliths, where there is much more variation in heterogeneity. Thus, it would be much easier to design a rover and other such necessary mining equipment to traverse the pyroclastic deposits. Mining and excavation will be straightforward in this material due to its fine-grained nature. These deposits have an excellent resource potential as they are relatively high in solar wind contents and the bulk chemistry of the glass is enriched in Si, Fe, Mg, and Ti. Also, these deposits would provide an extremely useful source

of shielding and construction material. The high density of the deposit materials makes these glass beads ideal for radiation protection. Also, the combination of high density and potentially thick, fine-grained deposits make tunnelling for habitats extremely attractive. A mere 10-20 m covering of this material will provide a year-round ambient temperature of zero degrees C and would almost eliminate radiation hazards.

Recent laboratory studies performed here at JSC by Allen *et al.* (1994) demonstrated that the Apollo 17 orange and black glasses release the most oxygen when exposed to hot hydrogen than any other lunar material. Compared to other resource feedstock sources on the Moon, these pyroclastic materials may just be the best oxygen resource on the Moon. Additionally, we now know that the orange and black glass spheres are rich in easily accessible sulfur and other sublimates as mentioned above. These samples have approximately 700 ppm by weight, sulfur in the bulk glass sample. Sulfur may be easily extracted from the lunar glass beads by heating them to relatively low temperatures compared to other samples. Gibson and Moore (1974) heated beads from sample 74220,84 and determined that sulfur is released at temperatures as low as 250°C to 650-700°C. The iron-rich pyroclastics may also be very good feedstocks for producing construction material such as sintered concrete-like blocks and for extracting metallic iron for a lunar base power system.

CONCLUSION

An evaluation of the remote sensing data confirm that lunar pyroclastic deposits are more prevalent and widespread than initially mapped. In addition, more material is thought to have erupted and been distributed than previously thought. Results of the geologic and remote sensing studies now lead us to believe that some pyroclastic deposits are exposed remnants of much larger regional deposits that are now largely buried by crater ejecta and/or subsequent lava flows.

Laboratory compositional evaluations and soil studies support the use of these materials as resource feedstock. Not only are they readily available in the most useful form - friable, fine-grained spheres, they are full of needed resource elements which can be readily beneficiated and used to establish and maintain a lunar base.

As with so many topics, we now know enough about lunar pyroclastics to know that we need to know much more before we can claim we understand them. Recently returned satellite data and future missions such as Questions we will address in the future with the help of the higher resolution Clementine and Galileo data include: What is the 'true' geographic extent of the pyroclastic deposits? Can we actually identify potential source vents and how do the volcanic deposits relate to them? How does the composition of a deposit vary? How does the Apollo 17/Taurus-Littrow deposit, a comparatively well studied site from which we have samples, relate to similar, remote sites like Sulpicius Gallus or Rima Bode? How well can we extrapolate our findings to other sites? How do the regional sites compare to the local sites? How do the glass samples in most lunar soils fit into the picture? We know that the Apollo 17/Taurus Littrow spectra and samples correlate, but how far can we extend this correlation and our comparative methods? Do Apollo 11, 12, 14, 15, and 16 samples correlate with Clementine/Galileo spectra, and if so, how? What do the fresh craters tell us about these deposits? Can we find fresh, pristine exposures on the crater rims/walls? Can we determine the relationship of regolith maturity to grain color? How do the spectral signatures of individual pyroclastic glass particles relate to the signatures of larger scale deposits? What is the relationship, if any, between the regional and local deposits?

REFERENCES

- Adams, J.B., Pieters, C.M., and McCord, T.B. 1974. Orange glass: Evidence for regional deposits of pyroclastic origin on the Moon. Proc. Lunar and Planet. Sci. Conf. 5th, pp. 171-186.
- AFGIT (Apollo Field Geology Investigation Team) 1973. Geologic exploration of Taurus-Littrow: Apollo 17 landing site. Science 182:672-680.
- Allen C.C, G.G. Bond, D.S. McKay (1994) Lunar oxygen production - a maturing technology, in Engineering, Construction and Operations in Space IV, pp. 1157-1166, Am. Soc. Civil Eng., New York, 1994.
- Andre, C.G., Hallam, M.E., Weidner, J.R., Podwysoki, M.H., Philpotts, J.A., Clark, P.E., and Adler, I. 1975. Correlation of Al/Si X-ray fluorescence data with other remote sensing data from the Taurus-Littrow area. Proc. Lunar and Planet. Sci. Conf. 5th, pp. 2739-2748.
- Bailey, N.G. and Ulrich, G.E. 1975. Apollo 17 voice transcript pertaining to the geology of the landing site. U.S.G.S. Rpt. 74-031, 361 pp.
- Bielefeld, M.J. 1977. Lunar surface chemistry of regions common to the orbital X-ray and gamma-ray experiments. Proc. Lunar Sci. Conf. 8th, pp. 1131-1147.
- Butler, P.Jr. 1978. Recognition of lunar glass droplets produced directly from endogenous melts: The evidence from S-Zn coatings (abstract). In Lunar and Planet. Sci. IX, (Lunar and Planetary Institute, Houston), pp. 143-145.
- Carr, M.H. and Meyer, C.E. 1972 Chemical and petrographic characteristics of the regolith at the Apollo 15 landing site. In The Apollo 15 Lunar Samples. (The Lunar Science Institute, Houston), pp. 48-50.
- Carter, J.L., Padovani E., and Taylor, H.C. 1973. Morphology and chemistry of particles from Apollo 17 soils 74220, 74241 and 74081 (abstract). EOS, Trans. Amer. Geophys. Union 54:582-584.
- Casella, J.C. and Binder, A.B. 1972. Geologic map of the Cleomedes Quadrangle of the Moon. U.S.G.S. Rpt. I-707, (LAC 44).
- Cavaretta, G., Funicelleo, R., Giles, H., Nicholls, G.D., Taddeucci, A., and Zussman, J. 1972. Geochemistry of green glass spherules from Apollo 15. In The Apollo 15 Lunar Samples. (Lunar Science Institute, Houston), pp. 202-203.
- Chao, E.C.T., Boreman, J.A., Minkin, J.A., James, O.B., and Desborough, G.A. 1970. Lunar glasses of impact origin: Physical and chemical characteristics and geologic implications. J. Geophys. Res. 75:7445-7479.
- Chen, H.K., Delano, J.W., and Lindsley, D.H. 1982. Chemistry and phase relations of VLT volcanic glasses from Apollo 14 and Apollo 17. Proc. Lunar Planet. Sci. Conf. 13th, in, J. Geophys. Res. 87:A171-181.
- Coombs, C.R. 1988. Explosive Volcanism on the Moon and the Development of Lunar Sinuous Rilles and Their Terrestrial Analogs. Ph.D. Dissertation, Univ. of Hawaii, 241 pp.
- Coombs, C.R. and Hawke, B.R. 1988. Explosive volcanism on the Moon: A review. Proc. Kagoshima Int'l Conf. on Volcanoes. Kagoshima, Japan. pp. 416-419.
- Coombs, C.R., Hawke, B.R., Lucey, P.G., Owensby, P.D., and Zisk, S.H. 1990. The Alphonsus region: A geologic and remote-sensing perspective. Proc. 20th Lunar and Planet. Sci. Conf., pp. 161-174.
- Coombs, C.R., McKay, D.S., Heiken, G.H., and Hawke, B.R. 1991. Should the future lunar base be located on a pyroclastic deposit? (abstract) Conf. on Near Earth Space Resources, Tuscon, AZ.
- Delano, J.D. 1986. Pristine lunar glasses: Criteria, data and implications. Proc. 16th Lunar and Planet. Sci. Conf. In J. Geophys. Res. 91:D201-D213.
- Delano, J.D. and Livi, K. 1981. Lunar volcanic glasses and their constraints on mare petrogenesis. Geochim. Cosmochim. Acta, 45: 2137-2149.
- Gaddis, L.R., Adams, J.B., Hawke, B.R., Head, J.W., McCord, T.B., Pieters, C.M., and Zisk, S.H. 1981. Characterization and distribution of pyroclastic units in the Rima Bode region of the Moon (abstract). In Lunar and Planet. Sci. Conf. XII (Lunar and Planet. Institute, Houston), pp. 318-320.
- Gaddis, L.R., Pieters, C.M., and Hawke, B.R. 1985. Remote sensing of lunar pyroclastic mantling deposits. Icarus 61:461-489.
- Gibson E.K. and G. W. Moore (1974) Sulfur abundances and distributions in the vally Taurus-Littrow. Proc. Lunar Sci. Conf. 5th, p. 1823-1837.

- Gibson, M.A. and Knudsen, C.W. 1985. Lunar oxygen production from ilmenite. In Lunar Bases and Space Activities of the 21st Century, ed. W.W. Mendell, (Lunar and Planetary Institutes, Houston). pp. 543-550.
- Glass, B.P. 1976. Major element composition of glasses from Apollo 11, 16 and 17 soil samples. Proc. Lunar Sci. Conf. 7th, 679-693.
- Green, D.H., Ringwood, A.E., Hibberson, W.O, and Ware, N.G. 1975. Experimental petrology of Apollo 17 mare basalts. Proc. Lunar Sci. Conf. 6th, 871-893.
- Hawke, B.R., Coombs, C.R., and Clark B. 1989a. Pyroclastic deposits: An ideal lunar resource (abstract). In Lunar and Planetary Science XX, (Lunar and Planetary Institute, Houston). pp. 387-390.
- Hawke, B.R., Coombs, C.R., and Clark B. 1989b. Ilmenite-rich pyroclastic deposits: An ideal lunar resource. Proc. Lunar and Planet. Sci. Conf. 20th, pp. 249-258.
- Hawke, B.R., Coombs, C.R., Gaddis, L.R., Lucey, P.G., and Owensby, P.D. 1989c. Remote sensing and geologic studies of localized dark mantle deposits on the Moon. Proc. Lunar and Planet. Sci. Conf. 19th, pp. 255-268.
- Hawke, B.R., Coombs, C.R., and Clark B. 1990. Pyroclastic deposits as sites for lunar bases. Proc. Space 90 Conf., pp. ??-??.
- Head, J.W. 1974. Lunar dark mantle deposits: Possible clues to the distribution of early mare deposits. Proc. Lunar and Planet. Sci. Conf. 5th, pp. 207-222.
- Head, J.W. and Wilson, L. 1979. Alphonsus-type dark halo craters: Morphology, morphometry and eruption conditions. Proc. Lunar and Planet. Sci. Conf. 10th, pp. 2861-2897.
- Heiken, G.H, McKay, D.S. , and Brown, R.W. 1974. Lunar deposits of possible pyroclastic origin. Geochim. Cosmochim. Acta 38:1703-1718.
- Huneke, J.C., Podeseck, F.A., and Wasserburg, G.J. 1973. An argon bouillabaisse including ages from the Luna 20 site (abstract). In Lunar Science IV, (The Lunar Science Institute, Houston), pp. 403-405.
- Hussein, L. 1972. The Ar-Ar and cosmic ray exposure ages of Apollo 15 crystalline rocks, breccias and glasses (abstract). In The Apollo 15 Lunar Samples, (The Lunar Science Institute, Houston), pp. 374-376.
- Kulcinski, G.L. 1988. ed., Astrofuel for the 21st Century. (College of Engineering, University of Wisconsin, Madison). 20 pp.
- Kulcinski, G.L., Santarius, J.F., and Wittenberg, L.J. 1986. Presentation at the 1st Lunar Development Symposium. September, 1986, Atlantic City, N.J.
- Lucey, P.G., Hawke, B.R., Pieters, C.M., Head, J.W., and McCord, T.B. 1986. A compositional study of the Aristarchus region of the Moon using near-infrared reflectance spectroscopy. Proc. Lunar and Planet. Sci. Conf. 16th in J. Geophys. Res. 91:D344-D354.
- Lucchita, B.K. 1973. Photogeology of the dark material in the Taurus-Littrow region of the Moon. Proc. Lunar Planet. Sci. Conf. 4th, pp. 149-162.
- Lucchita, B.K. and Schmitt, H.H. 1974. Orange material in the Sulpicius Gallus Formation at the southwestern edge of Mare Serenitatis. Proc. Lunar and Planet. Sci. Conf. 5th, pp. 223-234.
- Marvin, U.B. and Walker, D. 1978. Implications of a titanium-rich glass clod at Oceanic Procellarum. Am. Mineral. 63:924-929.
- McCord, T.B., Pieters, C.M., and Feierberg, M.A. 1976. Multispectral mapping of the lunar surface using ground-based telescopes. Icarus 29:1-34.
- McCord, T.B., Grabow, M.A., Feierberg, M.A., MacLaskey, D., and Pieters, C.M. 1979. Lunar multispectral maps: Part II of the lunar nearside. Icarus 37:1-28
- McCord, T.B., Clark, R.N., Hawke, B.R., McFadden, L.A., Owensby, P.D., Pieters, C.M., and Adams, J.B. 1981. Moon: Near-infrared spectral reflectance, a first good look. J. Geophys. Res. 86:10,883-10,892.
- McGetchin, T.R. and Head, J.W. 1972. Lunar cinder cones. Science 180:68-71.
- McKay, D.S., Clanton, U.S., and Ladle, G. 1973. Scanning electron microscope study of Apollo 15 green glass. Proc. 4th Lunar Sci. Conf., pp. 225-238.
- McKay, D.S. and G.H. Heiken 1973. Petrography and scanning electron microscope study of Apollo 17 orange and black glass (abstract). EOS. Trans. Amer. Geophys. Union 54:599-600.
- Mendell, W.W. 1985. ed., Lunar Bases and Space Activities of the 21st Century. (Lunar and Planetary Institute, Houston), 866 pp.
- Nava, D.F. 1974. Chemical compositions of some soils and rock types from the Apollo 15, 16, and 17 lunar sites. Proc. Lunar Sci. Conf. 5th, pp. 1087-1096.

- Olson, A.B. and Wilhelms, D.E. 1974. Geologic map of Mare Undarum Quadrangle of the Moon. U.S.G.S. Rpt. I-837 (LAC 62).
- Pieters, C.M., McCord, T.B., Zisk, S.H., and Adams, J.B. 1973. Lunar black spots and the nature of the Apollo 17 landing area. J. Geophys. Res. 78:5867-5875.
- Pieters, C.M., McCord, T.B., Charette, M.P., and Adams J.B. 1974. Lunar surface: Identification of the dark mantling material in the Apollo 17 soil samples. Science 183:1191-1194.
- Pohn, H.A. and Wildey, R.L. 1970. A photoelectric-photographic study of the normal albedo of the Moon. U.S. G.S. Prof. Paper 559-E, 20 pp.
- Quaide, W.L. 1973. Provenance of Apennine Front regolith materials (abstract). In Lunar Science IV, (The Lunar Science Institute, Houston), pp. 606-607.
- Reid, A.M., Warner, J., Ridley, W.I., Johnston, D.A., Harmon, R.S., Jakes, P. and Brown, R.W. 1972. The major element compositions of lunar rocks as inferred from glass compositions in the lunar soils. Proc. Lunar Sci. Conf. 3rd. 363-378.
- Reid, A.M., Lofgren G.E., Heiken G.H., Brown, R.W., and Moreland, G. 1973a. Apollo 17 orange glass, Apollo 15 green glass and Hawaiian lava fountain glass (abstract). EOS, Trans. Amer. Geophys. Union, 54:606-607.
- Reid, A.M., Warner, J.L., Ridley, W.I., and Brown, R.W. 1973b. Luna 0 soil: Abundance and composition of phases in the 45-125 μ m fraction. Geochim. Cosmochim. Acta, 37:1011-1030.
- Rhodes, J.M., Rogers, K.V., Shih, C., Bansal, B.M., Niquist, L.E., Wiesmann, H., and Hubbard, N.J. 1974. The relationships between geology and soil chemistry at the Apollo 17 landing site. Proc. Lunar Sci. Conf. 5th, pp. 1097-1117.
- Ringwood, A.E. and Kesson, S.E. 1976. A dynamic model for mare basalt petrogenesis. Proc. Lunar Sci. Conf. 7th. 1697-1722.
- Roedder, E. and Weiblen, P. 1973. Origin of orange glass spherules in Apollo 17 sample 74220 (abstract). EOS, Trans. Amer. Geophys. Union, 54:612-613.
- Rose, H.J., Jr., Cuttita, F., Berman, S., Brown, F.W., Carron, M.K., Christian, R.P., Dwornick, E.J., and Greenland, L.P. 1974. Chemical composition of rocks and soils at Taurus-Littrow. Proc. Lunar Sci. Conf. 5th, pp. 1119-1133.
- Schmitt, H.H. 1973. Apollo 17 report on the Valley of Taurus-Littrow. Science 182:681-690.
- Schonfeld, E. and Bielefield, M.J. 1978. Correlation of dark mantle deposits with high Mg/Al ratios. Proc. Lunar Planet. Sci. Conf. 9th, pp. 3037-3048.
- Simon, M.C. 1985. A parametric analysis of lunar oxygen production. In Lunar Bases and Space Activities of the 21st Century, ed. W.W. Mendell, (Lunar and Planetary Institutes, Houston), pp. 531-541.
- Thompson, T.W. 1979. A review of Earth-based radar mapping of the Moon. Moon and Planets 20:179-198.
- Warner, R.D., Taylor, G.J., and Keil K. 1979. Composition of glasses in Apollo 17 samples and their relation to known lunar rock types. Proc. Lunar Planet. Sci. Conf. 10th. 1437-1456.
- Whitaker, E.A. 1972. Lunar color boundaries and their relationship to topographic features: A preliminary survey. The Moon 4:348-355.
- Zisk, S.H., Pettengill, G.H., and Catuna G.W. 1974. High resolution radar map of the lunar surface at 3.8-cm wavelength. The Moon 10:17-50.

DESIGN OF PLANT GAS EXCHANGE EXPERIMENTS
IN A VARIABLE PRESSURE GROWTH CHAMBER

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1995

Johnson Space Center

Prepared By:	Kenneth A. Corey, Ph.D.
Academic Rank:	Associate Professor
University & Department	University of Massachusetts Dept. of Plant & Soil Sciences Amherst, MA 01003
NASA/JSC	
Directorate:	Engineering
Division:	Crew & Thermal Systems
Branch:	Life Support & Thermal Systems
JSC Colleague:	Daniel J. Barta, Ph.D.
Date Submitted:	August 1, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

Sustainable human presence in extreme environments such as lunar and martian bases will require bioregenerative components to human life support systems where plants are used for generation of oxygen, food, and water. Reduced atmospheric pressures will be used to minimize mass and engineering requirements. Few studies have assessed the metabolic and developmental responses of plants to reduced pressure and varied oxygen atmospheres. The first tests of hypobaric pressures on plant gas exchange and biomass production at the Johnson Space Center will be initiated in January 1996 in the Variable Pressure Growth Chamber (VPGC), a large, closed plant growth chamber rated for 10.2 psi. Experiments were designed and protocols detailed for two complete growouts each of lettuce and wheat to generate a general database for human life support requirements and to answer questions about plant growth processes in reduced pressure and varied oxygen environments.

The central objective of crop growth studies in the VPGC is to determine the influence of reduced pressure and reduced oxygen on the rates of photosynthesis, dark respiration, evapotranspiration and biomass production of lettuce and wheat. Due to the constraint of one experimental unit, internal controls, called pressure transients, will be used to evaluate rates of CO₂ uptake, O₂ evolution, and H₂O generation. Pressure transients will give interpretive power to the results of repeated growouts at both reduced and ambient pressures. Other experiments involve the generation of response functions to partial pressures of O₂ and CO₂ and to light intensity. Protocol for determining and calculating rates of gas exchange have been detailed. In order to build these databases and implement the necessary treatment combinations in short time periods, specific requirements for gas injections and removals have been defined. A set of system capability checks will include determination of leakage rates conducted prior to the actual crop growouts. Schedules of experimental events for lettuce and wheat are outlined and include replications in time of diurnal routines, pressure transients, variable pO₂, pO₂/pCO₂ ratio, and light intensity responses.

INTRODUCTION

Pressure and composition of atmospheres in future life support systems such as lunar and martian bases will likely differ from those of a sea-level, earth-based environment. Establishing an atmosphere for advanced closed life support systems involves consideration of requirements suitable for both autotrophs and heterotrophs. Human life support includes requirements for oxygen supply and carbon dioxide removal, roles served at least in part by plants in advanced life support systems having a bioregenerative component. In recent years, large, closed, controlled environment chambers have been constructed by NASA for gas exchange and plant growth studies (4, 10,22,29,31,32).

Optimum gas environments for plants and people may not be the same even though the atmosphere of the biosphere has evolved to establish gas compositions which reflect the reciprocal exchange of oxygen and carbon dioxide. While the atmospheric composition of the biosphere is suitable for the survival of both, it is not optimum for all growth and development processes. For example, it is well established that the carbon dioxide concentration of the biosphere is considerably lower than the level at which photosynthesis of plants is saturated (2,6,12,16,18). This knowledge has been applied to crop production in greenhouses and other closed environment settings where atmospheres are commonly enriched with carbon dioxide to accelerate growth rates.

Hypobaric Effects

Hypobaric pressures will likely be used to decrease the mass and engineering requirements for establishing and sustaining life support systems at extraterrestrial outposts. Recent studies suggest that plant growth may be enhanced at hypobaric pressures particularly when combined with decreased partial pressures of oxygen (1,8,9,11,19,23). At present, it is not clear what may explain effects of hypobaric pressure on plant processes, particularly photosynthesis. However, there are two major possibilities based on well established physical principles and the current state of knowledge in plant metabolism. First, the diffusion coefficient of gases in air increases at atmospheric pressures below those of ambient sea level. The relationship is expressed as

$$D' = D^0(T_1/293)^m(760/P_1),$$

where D^0 is the diffusion coefficient in air at 293 K and 760 mm Hg in cm^2/s , T is the temperature in K, P_1 is the pressure in mm Hg, and m is a factor that depends on characteristics of the diffusing gas and for CO_2 in air is equal to 2. The calculated diffusion coefficients in air for CO_2 and H_2O at a range of pressures at 23 C (baseline temperature to be used for crop production tests) are presented in Figure 1. Comparing ambient sea level pressure with the baseline pressure to be used for hypobaric plant growth experiments illustrates that the diffusion coefficients for CO_2 and H_2O increases by a factor of 1.44 at the lower pressure. Increased rates of CO_2 diffusion will result in a more rapid rate of transport to the site of photosynthesis assuming that stomatal conductance and leaf boundary layer resistance also remain unaffected (13). Photosynthetic enhancement attributable to increased carbon dioxide diffusivity would be expected to be greater at carbon dioxide concentrations well below saturation with the effect becoming negligible as

saturation concentrations are approached. Also, the CO₂ concentration required to saturate P_s is highly dependent on genotype (mesophyll resistance), light intensity, and temperature. Water diffusion at reduced pressures will also be enhanced and will lead to

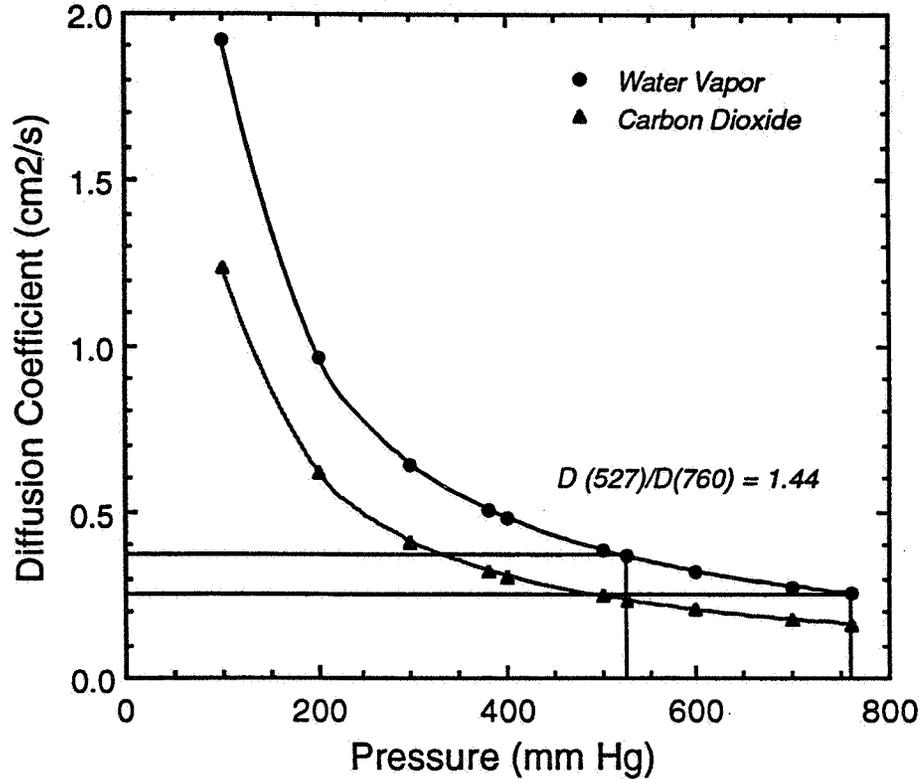


Figure 1.- Diffusion coefficients of CO₂ and H₂O calculated for a range of pressures at 23 C.

greater rates of transpiration (T_s). Increased rates of T_s with decreasing pressure have been reported in several studies (11,14,19). Given that the vapor pressure deficit (VPD) is held constant at different P_t, it should then be possible to determine the influence of P_t on T_s solely through its effect on D(H₂O^v).

A second major possibility for reduced pressure effects relates to the effects of O₂ on plant metabolic processes. If a reduction in atmospheric pressure is accompanied by a reduction in the partial pressure of oxygen, other effects may occur considering the range of oxygenases that regulate plant metabolism. A well established effect of decreased oxygen partial pressure is the decrease in activities of ribulose biphosphate carboxylase-oxygenase (RUBISCO) and glycolic acid oxidase (12,16). The net effect of the activities of RUBISCO and photorespiratory enzymes in C-3 plant species is an increased carbon loss which can be lessened at lowered partial pressures of oxygen (15,18,20,21,25,26).

Thus, optimizing productivity of plants also may involve a combination of hypobaric pressures and lowered partial pressures of oxygen. In addition, there is an effect of lowered oxygen on background or dark respiration (DR) (7,8,9,17,27,28) which may further increase the rate of biomass production through decreased carbon loss. Since cytochrome oxidase, the terminal electron acceptor in the mitochondrial electron transport chain, has a high affinity for oxygen and the stomatal resistance in the light is low, there may be only minimal effects on background respiration of shoots in the light at pO_2 down to about 1.5 psi. There may be greater effects of pO_2 on root respiration depending on the nature of the culture medium (e.g. solid matrix, solution culture, nutrient film technique), and the density and resistance of the root mass and root surfaces. For certain plant species such as wheat and lettuce, which can be grown under continuous light, the diminution of overall, whole canopy DR (shoots and roots) may increase the rate of crop growth. However, in some plants there may be intermediates produced at sea level pO_2 values which serve key roles in developmental processes such as seed germination. Perhaps there are different pO_2 optima for the periods before and after attainment of photosynthetic competence during seedling development.

Defining Atmospheres

Given that people and plants have different optima for atmospheric regimes, what atmospheres should be generated which would be suitable for an integrated system, and alternatively, for isolated systems? The issue of designing appropriate atmospheres for people in isolation has been well researched and various regimes have been implemented for space missions (Table 1). An open, but perhaps not critical issue in human physiology is whether there are any detrimental effects of prolonged exposure to lowered atmospheric pressures which are necessarily accompanied by lowered partial pressures of N_2 . At this point, there is no strong evidence to indicate that N_2 , the major component of the Earth's atmosphere, serves any vital long range function in human physiology.

A set of gas composition and atmospheric pressure values similar to those for human life support is not established for plants, though physiological responses to partial pressures of CO_2 and O_2 have been documented. Furthermore, since there is a greater genetic diversity to consider for plants than for the human species, responses of individual species and genotypes of plants to varied atmospheric conditions are expected to vary more widely than responses of humans. Approaches to the problem of defining atmospheres for plants could be tailored to the situation. One scenario is to define those conditions which are optimum for growth and development regardless of human needs. It is likely that such conditions may not also be suitable to people, necessitating compartment isolation and procedures to enable brief periods of human integration with a plant compartment. If isolation is feasible in future closed environments and if it translates into more efficient production of biomass, then implementation may be advantageous. Working from this assumption, how is the "best" atmospheric regime for plants determined? There are few reports on this subject except for the aforementioned effects of carbon dioxide enrichment and lowered partial pressure of oxygen. What effects, if any, will decreased atmospheric pressures and decreased partial pressures of oxygen have on rates of biomass production?

Another approach to defining atmospheres for plants is to assume a regime which will be integrated with people. This automatically places constraints on the minimum oxygen partial pressure and maximum carbon dioxide partial pressure which can be used safely and comfortably. The partial pressure of carbon dioxide which maximizes plant

TABLE 1.- ATMOSPHERIC REGIMES USED FOR HUMAN LIFE SUPPORT IN THE U.S. SPACE PROGRAM.^a

Program	Total (kPa)	Pressure ^b	
		Oxygen (kPa)	Carbon Dioxide (Pa)
Apollo	34.5 (5.0)	34.5 (5.0)	-----
Skylab	34.5 (5.0)	22.7 - 26.9 (3.3 -3.9)	660 (5.0)
Orbiter	101.2 (14.7)	20.3 - 23.8 (3.0 - 3.5)	1010 (7.6)
Orbiter (EVA prep)	70.3 (10.2)	18.3 - 19.3 (2.7 - 2.8)	1010 (7.6)
Space Station Freedom	70.3 (10.2)	68.9 - 71.7 (10.0 - 4.0)	1013 (7.6)

^aadapted from Tables B-1 and B-3 in Wieland, P.O., 1994, pages 184-185 (ref. 33).

^b psi given in parentheses except carbon dioxide where it is mm Hg.

growth is fortunately well below the defined safety limit for people. The partial pressure of oxygen is perhaps considerably higher than what could be established for optimum plant growth and would have to be a set condition based on human requirements. However, there is considerable latitude in the selection of the total pressure environment, which, for safety issues, would involve the use of a diluent/quenching gas such as nitrogen or helium. During the Skylab missions, the atmosphere consisted of an oxygen partial pressure of 160 mm Hg and 90 mm Hg nitrogen, with carbon dioxide maintained below about 5 mm Hg. The oxygen level was the same as ambient sea level, but the reduced nitrogen gave a total pressure environment of about one-third atmosphere with no documented deleterious effects on human health for a mission that lasted nearly three months. A reduced pressure database for plants needs to be developed which will include conditions that meet the goal of optimizing plant productivity and conditions which would also be suitable for human life support.

Variable Pressure Growth Chamber

The first tests of hypobaric pressures on plant gas exchange and biomass production at the Johnson Space Center will be initiated in January 1996 using the Variable Pressure Growth Chamber (VPGC) with tests of chamber and subsystem function beginning in September 1995. The VPGC, which is rated for a 10.2 psi pressure atmosphere and has a growth area capability of 11.2 m² (4) will be used for two complete growouts each of lettuce (*Lactuca sativa* 'Waldmann's Green') and wheat (*Triticum aestivum* 'Yecora Rojo' or a new dwarf genotype, 'Utah-20-1-41'). A primary objective of these experiments is to determine the effects of subambient pressure on gas exchange

and yield of plant species selected as candidate crops for the Controlled Ecological Life Support Systems (CELSS) program. The mass balance of metabolic gases will be important for matching plant requirements and resource generation to the resource needs of humans (3). A set of values for the key environmental variables affecting plant growth will be established and at specific stages of development, experiments will be conducted to answer specific questions about gas exchange responses under the subambient pressure environment. A peripheral, but important issue related to plant growth in closed environments is the evolution of volatiles which may accumulate to physiologically active levels. The plant volatile and hormone ethylene, is of special concern because of its wide range of metabolic effects. Since the synthesis and action of ethylene are dependent on the partial pressure of oxygen, ethylene concentration will be measured routinely and will be of particular interest in the context of variable pO_2 experiments.

Large scale experiments in growth chambers or in chambers designed as prototype CELSS have been concerned mainly with defining and implementing environmental conditions considered optimum for plant growth. Growouts of specific crops may also involve the conduct of experiments to characterize gas exchange responses to light, carbon dioxide, temperature, and vapor pressure deficit. The VPGC presents the opportunity to conduct similar experiments at subambient pressure. Unique to the scale of this chamber and a pertinent focus to experiments conducted during VPGC growouts will be the use of different partial pressures of oxygen.

OBJECTIVES

A. Database

The following objectives relate to the overall goal of developing a database on crop gas exchange and biomass production for advanced closed life support systems using the VPGC as a tool.

1. Determine gas exchange rates (i.e. P_s , DR , and E_t) of lettuce and wheat throughout growth and development.
2. Compare biomass production and rates of gas exchange in hypobaric and ambient pressure environments.
3. Develop photosynthetic response functions for light intensity (PPF) and carbon dioxide concentration.
4. Determine changes in concentration of ethylene in the atmosphere during crop development.

B. Investigative

The following is a list of questions relevant to issues of plant processes and growth in closed life support systems and can be addressed on a large scale plant canopy using the VPGC as an experimental unit.

1. Does hypobaric pressure enhance photosynthesis and biomass production of lettuce and wheat?
2. Are rates of gas exchange (i.e. P_s , DR , and E_t) influenced by total atmospheric pressure and by partial pressure of oxygen?
3. What are the physiological explanations for any observed effects on gas exchange? Attempts will be made to make distinctions at the crop canopy level between metabolic and physical possibilities.
4. Are there interactive effects of partial pressures of oxygen and carbon dioxide on gas exchange rates, i.e. is net photosynthesis a function of the pO_2/pCO_2 ratio?

EXPERIMENTAL PROTOCOLS

Baseline Conditions

Baseline environmental conditions were selected based on previous reports of experiments in plant growth chambers (5,6,10,24,31,32) and on the unique capabilities of the VPGC for maintaining reduced atmospheric pressure (Table 2).

TABLE 2.- BASELINE CONDITIONS FOR CROP PRODUCTION TESTS IN VPGC.

Environmental Condition	Lettuce	Wheat
Pressure (psi) ^a	10.2	10.2
Oxygen (psi) ^b	2.1	2.1
Carbon dioxide ($\mu\text{mol/mol}$)	1200	1200
Temperature (C)	23	23
Dewpoint temperature (C) ^c	17.5	17.5
Photoperiod (light/dark)	18/6	20/4
PPF ($\mu\text{mol/m}^2/\text{s}$)	400	1500

^aPressure given is for first growout of each crop; the second growout will use 14.7 psi as the baseline.

^bPartial pressure of O₂ given is for first growout with the second growout to be set at 3.1 psi.

^cA dewpoint temperature was selected to give a relative humidity of 70%.

The following sections are the categories of experiments and types of data to be compiled for the two growouts of lettuce and wheat.

Diurnal Routines

Each day, measurements of net photosynthesis (Ps), dark respiration (DR), and evapotranspiration (Et) will be made and continued throughout growth and development. Sensitive measurements of rates of CO₂ uptake will probably not be possible until following the completion of the germination process (several days after planting). Since there will be light/dark cycles used and no attempt at CO₂ removal will be made, the DR measurement will be obtained from the slope of the linear increase in CO₂ concentration that occurs following lamp shutdown. It is assumed that DR is constant over the range of CO₂ concentrations that develop. After the lights are turned on the CO₂ will decrease linearly until the setpoint value of 1200 ppm is achieved. This is referred to as the diurnal photosynthetic drawdown. While the Ps drawdown method of measurement is not steady state, the baseline concentration of 1200 $\mu\text{mol/mol}$ is close to saturation of Ps at the

specified light intensity, is linear, and is therefore a reliable daily rate measurement (unpublished data). A supplementary method for measuring the photosynthetic rate will be achieved for the complete light period by tracking the absolute quantity of CO₂ injected to maintain the setpoint. This is equivalent to the amount of CO₂ used in Ps plus the quantity leaked from the chamber during the time of measurement. The numbers obtained are referred to as CO₂ mass flow measurements and are considered a steady state method of measurement. Leakage tests performed prior to and following each crop test will be used to apply corrections to calculations made for any long term measurements (i.e. those exceeding a couple hours). It is anticipated that the leak rate, particularly at the subambient pressures, will be small enough (i.e. < 5 %/day) to assume a negligible loss for short term (<1 hour) rate calculations. Oxygen concentration during the light phase will be maintained at the established setpoint and quantities tracked by continual scrubbing with a molecular sieve. Accounting of the oxygen removed during the photoperiod by the molecular sieve will also serve as a supplementary method of Ps rate measurement and will be compared with the CO₂ mass flow data. It may be possible to use the two measurements to calculate an assimilation quotient for Ps (CO₂ uptake/O₂ evolved).

Daily measurements of water flux will also be made for both light and dark components of the daily cycle. The light measurement actually represents a combined evapotranspiration (Et) value. It will not be possible to separate the two components precisely since the energy budgets differ and a finite, but perhaps small stomatal conductance can be measured during the dark period. An estimate of evaporation will be obtained in the pretest checkouts by running the system at test conditions without plants.

P_t Transients

Since the baseline pressure used is 10.2 psi, it will be necessary to compare measurements of gas exchange rates with those conducted at ambient pressure (see objectives A.2 and B.1). The constraint of one experimental unit and limited replications in time pose an interpretive challenge. Therefore, short-term rate measurements at several stages of development will be made. For the first growout of each crop, the pressure transients will be ambient pressure and for the second growout of each crop where the baseline atmospheric pressure is ambient, the pressure transients will be conducted at 10.2 psi. It will also be necessary to vary the partial pressure of O₂ during the pressure transient experiments. Both the reduced and normal sea level ambient partial pressures of O₂ will be used. These short term experiments will be used as an aid to interpretation of biomass production results and for establishing the extent to which overall rates of plant metabolism can be altered by P_t and pO₂. Pure N₂ and O₂ will be injected to establish increased partial pressures and total atmospheric pressures. The vacuum pumps and a method for O₂ removal (molecular sieve) will be used to reduce levels back to baseline or to change O₂ (for variable O₂ experiments). The pressure transients may be summarized as follows: Growout I (baseline P_t=10.2 psi) - P_t=14.7 psi, pO₂=3.1 psi and P_t=14.7 psi, pO₂=2.1, and Growout II (baseline P_t=14.7 psi) - P_t=10.2 psi, pO₂=2.1 psi and P_t=10.2 psi, pO₂=3.1psi.

Variable pO₂

Crucial to defining the optimum atmospheric environment for plant growth is the

determination of partial pressures of O₂ that will maximize P_s and minimize background respiration (DR). An experiment with variable partial pressure of O₂ will be conducted over the range of 1.1 psi to 3.1 psi using 5 levels for the light period measurements. Due to the constraint of a short dark period, only 2 levels of pO₂ will be used during the dark period. Within each experiment, the appropriate ambient pressure treatments (like those mentioned in the P_t transients section) will be used as internal controls. A sample protocol for the P_s measurements is illustrated in Figure 2; the exact order of the 8 separate measurement periods randomized for each run of the same experiment. The zero-slope portions of the figure represent the quasi-steady state measurements of CO₂ mass flow injections. A fifteen-minute period for total and partial pressure changes is shown and is close to the time limit constraint for changes necessary to complete the set of treatments. Mass flow of CO₂ will be tracked at intervals no greater than 5 minutes in order to establish sufficient time segments for strong statistical estimates of rates. For the dark period experiment, CO₂ will increase and then the starting value of 1200 ppm reestablished using a CO₂ removal system (LiOH) or a photosynthetic drawdown prior to beginning the next treatment. A sample protocol and example of a hypothetical result are shown in Figure 3 to

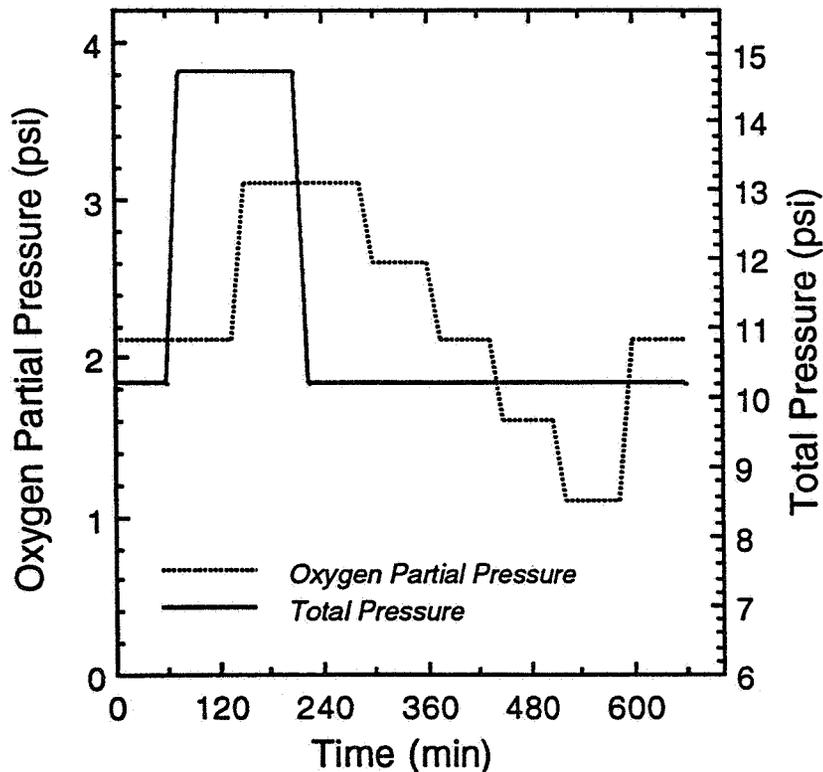


Figure 2.- Example of treatment protocol for the variable oxygen experiments in the VPGC during the light period.

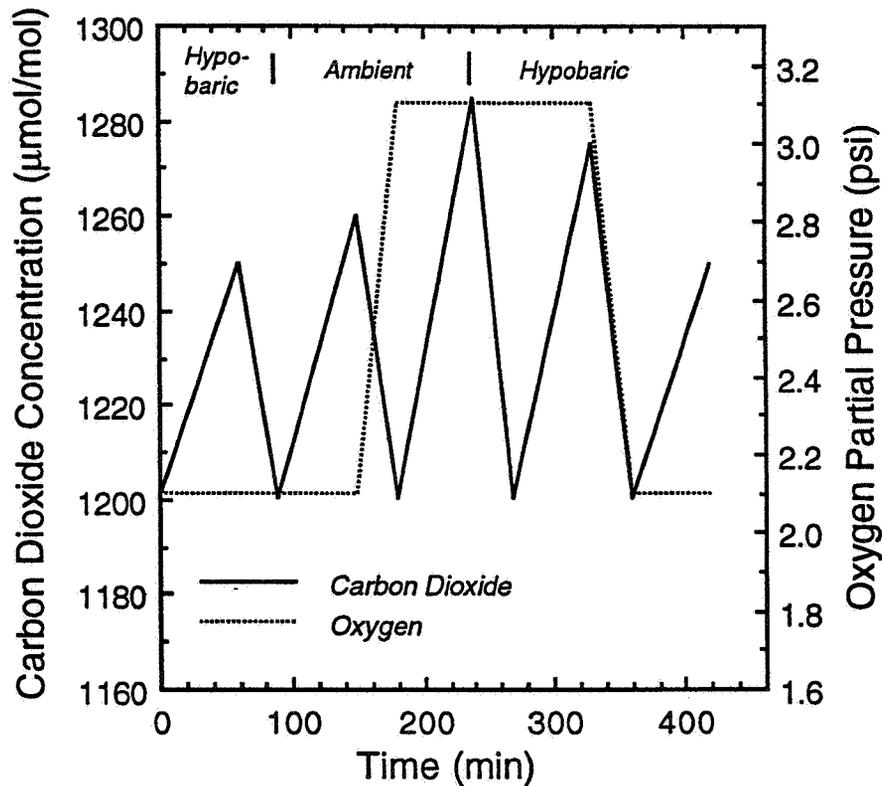


Figure 3.- Sample protocol and example of hypothetical results to illustrate experiments for measuring dark respiration during application of different treatment atmospheres.

illustrate the treatment sequence and possible effects. Also, tracking of water flux for both light and dark experiments would be desirable to determine if the treatments influence evapotranspiration. Given that water vapor pressure deficit is held constant, any treatment effects on E_t will likely be attributable to effects on stomatal conductance.

pO_2/pCO_2 Ratio

Possible pO_2 effects on the rate of P_s may actually be attributable to the ratio of oxygen to carbon dioxide. This experiment will enable the generation of photosynthetic response functions to CO_2 concentration at two levels of pO_2 (2.1 and 3.1 psi). Control of CO_2 will be disabled and the CO_2 allowed to decrease (photosynthetic drawdown). Drawdowns to the CO_2 compensation point, while desirable for generating a complete function, can take several hours and may result in a substantial decrease in the free carbohydrate pool. The latter effect may influence how the crop canopy is in turn affected

by a subsequent change in the pO_2 . The photosynthetic response curve can be characterized down to about 200 ppm CO_2 fairly quickly (i.e. 60-90 minutes) without a major change in the free carbohydrate pool. The method just described is nonsteady state and may not be representative of the internal conditions and time necessary for true photosynthetic steady state to be achieved. Stomatal conductance is strongly dependent on CO_2 concentration and the VPGC has a rather tight configuration (i.e. volume to area ratio = 2.4). Therefore, a steady state experiment will be conducted and mass flow measurements used to calculate the rate of P_s at different CO_2 concentration setpoints. If the two methods do not provide consistent results, then a more time-consuming steady state method (CO_2 setpoints and mass flow measurements) will be used.

Light Intensity Response Function

A routine experiment in constructing gas exchange databases for plants is to determine the photosynthetic response to light intensity (PPF) in the photosynthetically active radiation (PAR) waveband (400 - 700 nm). For lettuce, a reduced baseline light intensity is necessary to minimize the risk of tipburn, a condition exacerbated by controlled environment growth using solution culture or nutrient film techniques. However, given the potential responses to low pressure and decreased partial pressure of oxygen and the potential for accelerating O_2 generation or CO_2 removal, it will be of interest to determine the photosynthetic response to light intensity for differing atmospheric conditions. This can be achieved by conducting an experiment using at least 2 light intensities (baseline and a higher value) in combination with the 2 P_t levels and 2 pO_2 levels as described in the protocol for the variable oxygen experiments. Some exploratory work will be necessary to determine appropriate PPF values to use in combination with the atmospheric treatment variables. The procedure will involve short duration (30 min) drawdowns of CO_2 and calculation of the rate of P_s from the slopes of the line segments (10). For wheat, a photosynthetic response is anticipated throughout the light range capability of up to 1500 $\mu mol/m^2 \cdot s$. This response function will be generated by dimming the lamps to achieve PPF values in the range of 250 to 1500 $\mu mol/m^2 \cdot s$ in steps of approximately 250. The light compensation point will be calculated as the x-intercept from the fitted function. As with the lettuce, an experiment will be conducted to determine if light saturation varies with P_t and pO_2 .

A schedule of experimental events for each crop and the estimated times in the growth cycle at which they will occur are summarized in Tables 3 and 4.

TEST REQUIREMENTS

The conduct of experiments outlined and discussed in this report require changes in total pressure and in partial pressures of individual gases in the atmosphere. The required changes must be accomplished in certain time periods in order to impose all necessary atmospheric treatment regimes. The types of changes, methods of implementing the changes, and method capabilities are summarized in Table 5.

Leak Rate Tests

Experiments outlined in this report involve the determination of steady-state and

nonsteady-state rates of consumption and evolution of metabolic gases in a closed environment. These measurements are based upon changes in concentration over time or upon quantities injected to maintain a constant concentration. For time increments more than a couple hours or for measurements integrated over days, calculations will require a knowledge of the rate of leakage of the gas of interest from or into the system. Leakage rates of CO₂ and C₂H₄ from the VPGC will be determined with the chamber in full test mode with the exception of plants. Individual gases will be injected and samples taken from the inside air and the external ambient atmosphere to determine the concentration gradient. The leak rate may be calculated from the formula,

$$L = 1/t_2 - t_1 \times \ln [C_1 - C_{amb}/C_2 - C_{amb}],$$

where L is the leak rate in %/day, t is time, C₁ is the concentration of the gas at t₂, C₁ the concentration of the gas at t₁, and C_{amb} the concentration of the gas in the ambient atmosphere (30). In conducting the test, sufficient time should be allowed after injecting CO₂ and C₂H₄ for equilibration with the chamber environment as the gases will dissolve in water, the seals, and perhaps other materials.

REFERENCES

1. Andre, M., and D. Massimino. 1992. Growth of plants at reduced pressures: experiments in wheat-technological advantages and constraints. *Adv. Space Res.* 12: 97-106.
2. Acock, B., V.R. Reddy, H.F. Hodges, D.N. Baker, J.M. McKinion. 1985. Photosynthetic response of soybean canopies to full-season carbon dioxide enrichment. *Agron. J.* 77: 942-947.
3. Barta, D.J., J.S. Dominick, and M. Kallberg. 1995. Early human testing of advanced life support systems, phase I. Society of Automotive Engineers Tech. Paper 95-1490, July 10-14, 1995.
4. Barta, D.J. and D.L. Henninger. 1995. Johnson Space Center's regenerative life support systems test bed. *Adv. Space Res.* (In press).
5. Bugbee, B.G. 1990. Exploring the limits of crop productivity: a model to evaluate progress. Pages 1-23 in R.D. MacElroy, ed. *Controlled Ecological Life Support Systems: CELSS '89 Workshop*. NASA technical memorandum 102277, Moffett Field, CA.
6. Bugbee, B.G. and F.B. Salisbury. 1988. Exploring the limits of crop productivity. *Plant Physiol.* 88: 869-878.
7. Chun, C. and T. Takakura. 1993. Control of root environment for hydroponic lettuce production - rate of root respiration under various dissolved oxygen concentrations. American Society of Agricultural Engineering Paper No. 934040.
8. Corey, K.A., M.E. Bates, and S.L. Adams. 1995. Carbon dioxide exchange of lettuce plants under hypobaric conditions. *Adv. Space Res.* (In press).
9. Corey, K.A. 1994. Carbon dioxide exchange of lettuce in decreased pressure atmospheres and different diluent gases. NASA/ASEE Summer Faculty Fellowship Final Report.

10. Corey, K.A. and R.M. Wheeler. 1992. Gas exchange in NASA's biomass production chamber: A preprototype closed human life support system. *BioScience* 42:503-509.
11. Daunicht, H.J. and H.J. Brinkjans. 1992. Gas exchange and growth of plants under reduced air pressure. *Adv. Space Res.* 12: 107-114.
12. Ehleringer, J.R. 1979. Photosynthesis and photorespiration: Biochemistry, physiology, and ecological implications. *HortScience* 14: 217-222.
13. Gale, J. 1972. Availability of carbon dioxide for photosynthesis at high altitudes: theoretical considerations. *Ecology* 53: 494-497.
14. Gale, J. 1973. Experimental evidence for the effect of barometric pressure on photosynthesis and transpiration. In: *Plant Responses to Climatic Factors*, Proceedings of the Uppsala Symposium, UNESCO, Paris, pp. 289-294.
15. Gerbaud, A. and M. Andre. 1989. Photosynthesis and photorespiration in whole plants of wheat. *Plant Physiol.* 89: 61-68.
16. Govindjee, ed., 1983. *Photosynthesis, Vol II: Development, Carbon Metabolism and Plant Productivity*. Academic Press, Inc., New York.
17. Musgrave, M.E., W.A. Gerth, H.W. Scheld, and B.R. Strain. 1988. Growth and mitochondrial respiration of mungbeans (*Phaseolus aureus* Roxb.) germinated at low pressure. *Plant Physiol.* 86: 19-22.
18. Musgrave, M.E. and B.R. Strain. 1988. Response of two wheat cultivars to CO₂ enrichment under subambient oxygen conditions. *Plant Physiol.* 87: 346-350.
19. Ohta, H., E. Goto, T. Takakura, F. Takagi, Y. Hirokawa, and K. Takagi. 1993. Measurement of photosynthetic and transpiration rates under low total pressures. American Society of Agricultural Engineering Paper No. 934009.
20. Parkinson, K.J., H.L. Penman, and E.B. Tregunna. 1974. Growth of plants in different oxygen concentrations. *J. Expt. Bot.* 25: 135-145.
21. Poskuta, J. 1968. Photosynthesis, photorespiration, and respiration of detached spruce twigs as influenced by oxygen concentration and light intensity. *Physiol. Plant.* 21: 1129-1136.
22. Prince, R.P., and W.M. Knott, J.C. Sager, and S.E. Hilding. 1987. Design and performance of the KSC Biomass Production Chamber, Society of Automotive Engineers Technical Paper Series No. 871437, Warrendale, PA.
23. Rule, D.E. and G.L. Staby. 1981. Growth of tomato seedlings at sub-atmospheric pressures. *HortScience* 16: 331-332.
24. Salisbury, F.B., and B.G. Bugbee. 1988. Plant productivity in controlled environments. *HortScience* 23: 293-299.
25. Siegel, S.M. 1961. Effects of reduced oxygen tension on vascular plants. *Physiol. Plant.* 14: 554-557.
26. Taylor, D.L. 1942. Influence of oxygen tension on respiration, fermentation and growth in wheat and rice. *Amer. J. Bot.* 29:721-738.
27. Thimann, K.V. 1954. The physiology of growth in plant tissues. *Am. Scientist* 42:589-606.
28. Thornley, J.H.M. 1970. Respiration, growth and maintenance in plants. *Nature* 227: 304-305.
29. Wheeler, R.M. 1992. Gas exchange measurements using a large, closed plant growth chamber. *HortScience* 27:777-780.
30. Wheeler, R.M., J.H. Drese, and J.C. Sager. 1991. Atmospheric leakage and condensate production in NASA's Biomass Production Chamber. Effect of diurnal

- temperature cycles. John F. Kennedy Space Center. NASA Tech. Mem. 103819.
31. Wheeler, R.M., K.A. Corey, J.C. Sager, and W.M. Knott. 1993. Gas exchange characteristics of wheat stands grown in a closed, controlled environment. *Crop Science* 33:161-168.
 32. Wheeler, R.M., C.L. Mackowiak, J.C. Sager, N.C. Yorio, W.M. Knott, and W.L. Berry. 1994. Growth and gas exchange by lettuce stands in a closed, controlled environment. *J. Amer. Soc. Hort. Sci.* 119: 610-615.
 33. Wieland, P.O. 1994. *Designing for Human Presence in Space: An Introduction to Environmental Control and Life Support Systems.* NASA RP - 1324, 348 pages.

ACRONYMS, ABBREVIATIONS, & SYMBOLS

CELSS	Controlled Ecological Life Support Systems
D(CO ₂)	Diffusion coefficient of carbon dioxide
D(H ₂ O ^v)	Diffusion coefficient of water vapor
DR	Dark Respiration
Et	Evapotranspiration
Hb	Hypobaric
Pr	Photorespiration
Ps	Photosynthesis (net)
P _t	total pressure of atmosphere
pCO ₂	partial pressure of carbon dioxide
pN ₂	partial pressure of nitrogen
pO ₂	partial pressure of oxygen
PPF	photosynthetic photon flux
Ts	Transpiration
VPD	vapor pressure deficit
VPGC	Variable Pressure Growth Chamber

TABLE 3.- SCHEDULE OF TEST ACTIVITIES FOR GAS EXCHANGE EXPERIMENTS WITH LETTUCE IN VPGC.

Test Day	Activities/Experiments
1-8	Germination and seedling establishment/dark respiration measurements ^a
9	Gas exchange measurements ^b
10-14	Stand establishment/Daily routines ^c
15	Daily routines/Pressure transient experiment (dark)
16	Pressure transient experiment (light)
17	Variable pO ₂ experiment
18	Ps response to CO ₂ at 2 pO ₂ (drawdowns) or pO ₂ :pCO ₂ ratio experiment
19	PPF experiment at 2 pO ₂
20-21	Daily routine
22	Daily routines/Pressure transient experiment (dark)
23	Pressure transient experiment (light)
24	Variable pO ₂ experiment
25	Ps response to CO ₂ at 2 pO ₂ (drawdowns) or pO ₂ :pCO ₂ ratio experiment
26	PPF experiment at 2 pO ₂
27	Daily routine
28	Daily routine/Return to ambient P _t - measure rates/Harvest ^d

^aRespiration rate measurements during germination will be possible by about day 2 or 3.

^bEstimate of day for obtaining reliable Ps and Et measurements.

^cDaily routine includes measurements of Ps, DR, and Et at baseline conditions.

^dDay of harvest is estimate.

TABLE 4.- SCHEDULE OF TEST ACTIVITIES FOR GAS EXCHANGE EXPERIMENTS WITH WHEAT IN VPGC.

Test Day	Activity/Experiment
1- 4	Germination and seedling establishment/Dark respiration measurements ^a
5	Gas exchange measurements ^b
6-14	Stand establishment/Daily routines ^c
15	Daily routines/Pressure transient experiment (dark)
16	Pressure transient experiment (light)
17	Variable pO ₂ experiment
18	Ps response to CO ₂ at 2 pO ₂ (drawdowns) or pO ₂ :pCO ₂ ratio experiment
19	PPF experiment at 2 pO ₂
20-21	Daily routines
22	Daily routines/Pressure transient experiment (dark)
23	Pressure transient experiment (light)
24	Variable pO ₂ experiment
25	Ps response to CO ₂ at 2 pO ₂ (drawdowns) or pO ₂ :pCO ₂ ratio experiment
26	PPF experiment at 2 pO ₂
27-35	Daily routines ^d
36-40	Repeat of experiments conducted days 15-19 and days 22-26
41-42	Daily routines
43-47	Repeat of experiments conducted days 36-40
48-64	Daily routines ^e /Return to ambient P _t - measure rates/Harvest ^f

^aRespiration rate measurements during germination will be possible by about day 2 or 3.

^bEstimate of day for obtaining reliable Ps and Et measurements.

- ^cDaily routine includes measurements of Ps, DR, and Et at baseline conditions.
- ^dFlowering and seed set will take place during this interval and atmospheric manipulations will be minimized.
- ^eLong duration of daily routines at baseline conditions occurs during the latter stages of growth and development which includes seed fill and senescence.
- ^fDay of harvest is estimate.

TABLE 5.- TEST REQUIREMENTS TO IMPLEMENT ATMOSPHERIC CHANGES IN THE VPGC DURING PLANT GROWTH STUDIES.

Atmospheric Change	Direction of Change	Change		Method	Method Capability
		Minimum	Maximum		
P _t	increase	4.5 psi	4.5 psi	injection	^a 1000 l/min
	decrease	4.5 psi	4.5 psi	vacuum	^b 2000 l/s
pO ₂	increase	0.5 psi	2.0 psi	injection	^a 1000 l/min
	decrease	0.5 psi	0.5 psi	Rs vacuum & injection	variable ^b 2000 l/s ^a 1000 l/min
pCO ₂	increase	maintenance 0 - 10 ppm	1000 ppm	molecular sieve injection	^c 2 l/min ^d 3 - 30000 cm ³ /min
	decrease	0 - 100 ppm	1000 ppm	Ps LiOH	variable ^e 74 mmol/min
pN ₂	increase	0.5 psi	4.5 psi	injection	^a 1000 l/min
	decrease	0.5 psi	4.5 psi	vacuum	^b 2000 l/s
pH ₂ O ^v	increase	maintenance		mist injection	
	decrease	maintenance		heat exchanger cold H ₂ O coils	

^aTescom Corp. dome-loaded pressure regulators, wide range of capacities available.

^bapproximate requirement of tests is only 20 l/s based on an approximate chamber volume of 27,000 liters at STP.

^ceffective only for continuous scrubbing to maintain pO₂.

^dBrooks Instrument Mass Flow Controller Model 5850i.

^eMaximum capability needed, system specific method may need to be developed.

Investigation of Two-Phase Flows in Piping Bends and Elbows

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1995

Johnson Space Center

**Prepared By: Allen B. Duncan, Ph.D., Assistant Professor
Vincent M. Sciascia, Graduate Research Assistant**

**University and Department: University of Illinois at Chicago
Department of Mechanical Engineering
M/C 251, Room 2039 ERF
842 W. Taylor St.
Chicago, IL 60607-7022**

NASA/JSC

Directorate: Engineering

Division: Crew and Thermal Systems

Branch: EC3

JSC Colleague: Kathryn M. Hurlbert

Date Submitted: August 9, 1995

Contract Number: NGT-44-001-800

Abstract

An experimental investigation of the hydrodynamic characteristics of two-phase R-113 flow has been carried out. Straight tube pressure drop data, as a function of mass flow rate (mass flux) and flow quality has been obtained using the Two-Phase Flow Test Facility located in the Advanced Thermal Laboratories of the Crew and Thermal Systems Division at the Lyndon B. Johnson Space Center. Additionally, after successfully obtaining the straight tube pressure drop data, the test facility was modified in order to obtain pressure drop data for the flow of two-phase R-113 through 180° piping bends. Inherent instabilities of the test facility prevented the successful acquisition of pressure drop data through the piping bends.

The experimental straight tube data will be presented and compared with existing predictive correlations in an attempt to gain insight into the utility of such correlations as the basis for developing design criteria. A discussion of the instabilities which rendered successful acquisition of the piping bend data will be presented and suggestions will be made for eliminating these system tendencies. Finally, recommendations for future investigations, based on successful reconfiguration of the test facility, will be made.

Introduction

Power requirements for spacecraft and satellites continue to increase, necessitating similar increases in the effectiveness and efficiencies of spacecraft thermal management systems. Active two-phase cooling loops have been identified as systems which may satisfy the requirements of future spaceflight thermal control systems. However, archival literature documenting both experimental and analytical investigations of the pressure gradient characteristics of two-phase flow through piping bends and elbows, such as those which will be included in realistic system designs, is virtually non-existent. The current investigation will obtain baseline data for the two-phase flow test facility developed by the Crew and Thermal Systems Division. Additionally, the investigation will obtain data which will help characterize the pressure drop characteristics of two-phase flows through 180° piping bends. This investigation will compliment and support the TEEM flight experiment scheduled for launch in 1997.

During the summer period, an experimental investigation of the hydrodynamic characteristics of two-phase flows similar to those which would occur in the two-phase active thermal control systems (ATCS) of a space vehicle was undertaken. This investigation represents the first step of a long term effort which will develop mathematical models and experimental data for the prediction of the pressure drops which may be expected for two-phase flow through realistic systems which include piping elbows, expanding and contracting flow sections, and commonly utilized devices such as quick-disconnect fittings.

Experimental Investigation

Straight Tube Investigation

A schematic of the two-phase flow test facility utilized for the straight-tube flow characterization of the current investigation is illustrated in Fig. 1. The R-113 entered the positive displacement pump as single-phase liquid. The single-phase liquid was pumped through the evaporator section which consisted of a circulation heater where heat input could induce partial vaporization of the flowing R-113. The liquid or two-phase mixture which exited the evaporator section could be observed in the flow visualization section in order to characterize the nature of the two-phase or single-phase flow. Downstream of the visualization section, the fluid passed through the pressure drop test section. This pressure drop test section allowed the operator to choose either a low range (Baratron) or high range (Validyne) pressure transducer for the measurement of the pressure drop in the test section. The parameters which were directly varied in the experimental investigation were those of pump power input (controlling mass flow rate or mass flux) and evaporator input power (controlling flow quality). In this manner, the pressure drop in the straight-tube section was measured as a function of both mass flow rate (mass flux) and flow quality. The two-phase test facility was used in conjunction with a PC based data acquisition system for the recording and processing of each measured value obtained from system instrumentation.

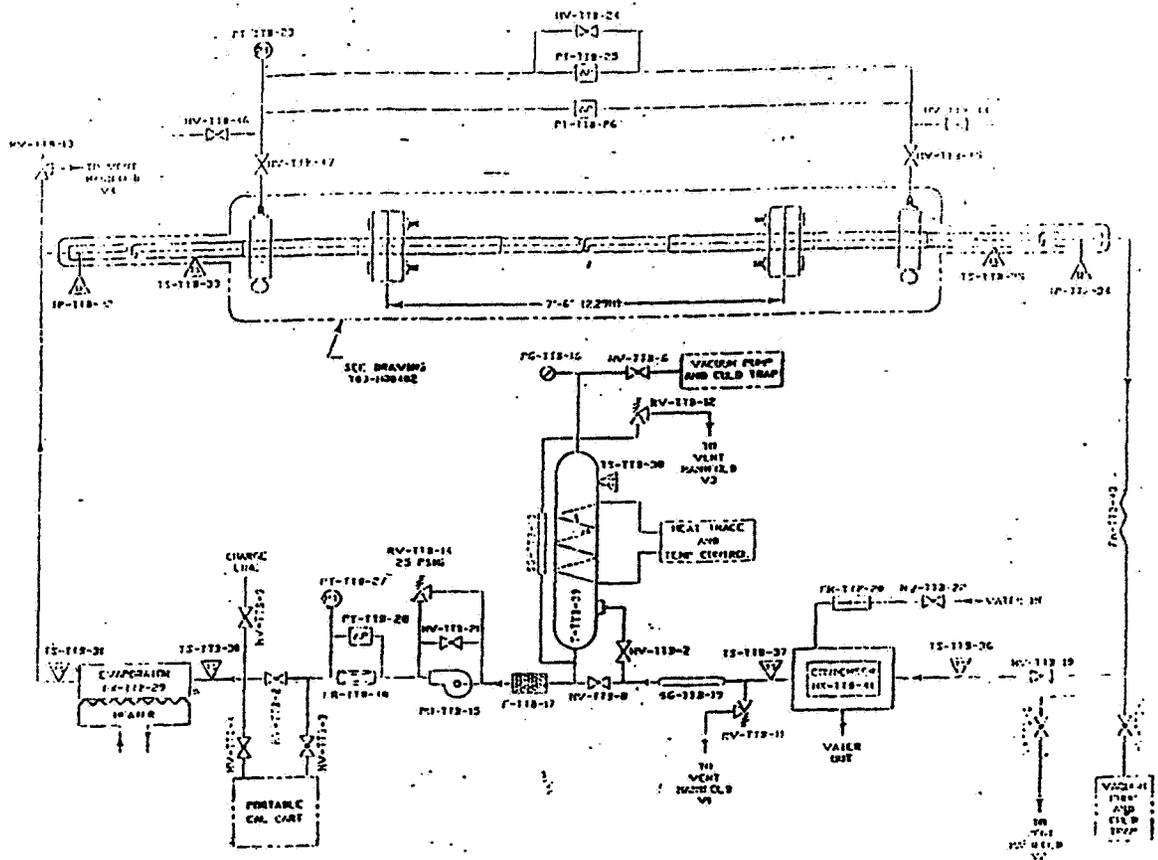


Fig. 1 Schematic of Test Facility

Figure 2 illustrates the measured pressure drop in the straight tube section for single phase R-113 liquid. While the measured values of pressure drop consistently exceed those which are predicted using the Blasius Solution¹, the trend exhibited was quite uniform, with the mean of the measure values exceeding the prediction by approximately 0.075 kPa for virtually the entire range of flow rates. Possible reasons for this steady state error are the presence of roughness in the tubes or a pressure drop due to the fittings present at each pressure tap and each end of the tube. This data indicated that for single phase flow, the measured values obtained from the test facility were acceptably accurate and repeatable.

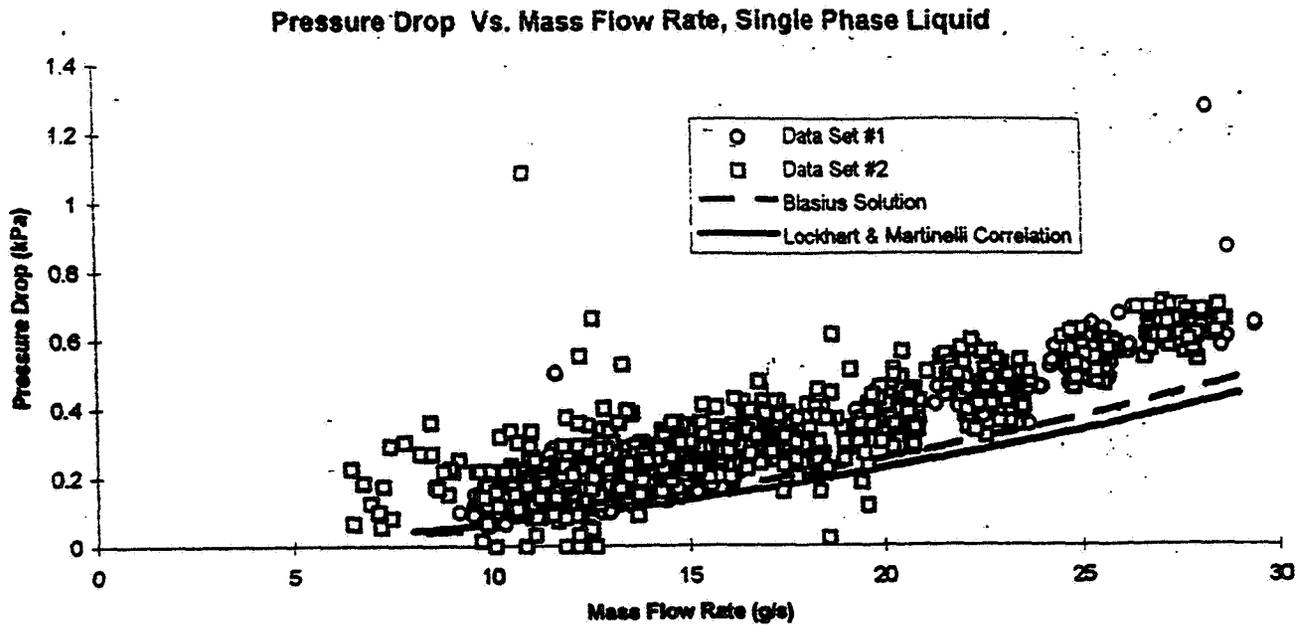


Fig. 2 Single Phase Straight Tube Pressure Drop, Data and Predictive Correlations

Figure 3 illustrates the measured pressure drop in the straight tube section for two-phase R-113 flow subjected to evaporator input levels which produced a flow quality of 20%. The data was compared with four predictive correlations, each of these empirically developed. The first of these was that of Lockhart and Martinelli², utilizing the friction factor associated with the Blasius Solution¹. Secondly, the prediction of Lockhart and Martinelli² was used with their recommended friction factor. The third predictive correlation utilized for comparison was that of Troniewski and Ulbrich³. Finally, the predictive correlation of Barozcy⁴ was utilized. While the trend of the observed data was consistent with those of the predictive correlations throughout the range of mass flow rates, each of the three correlations predicted pressure drop values which were greater than those which were measured.

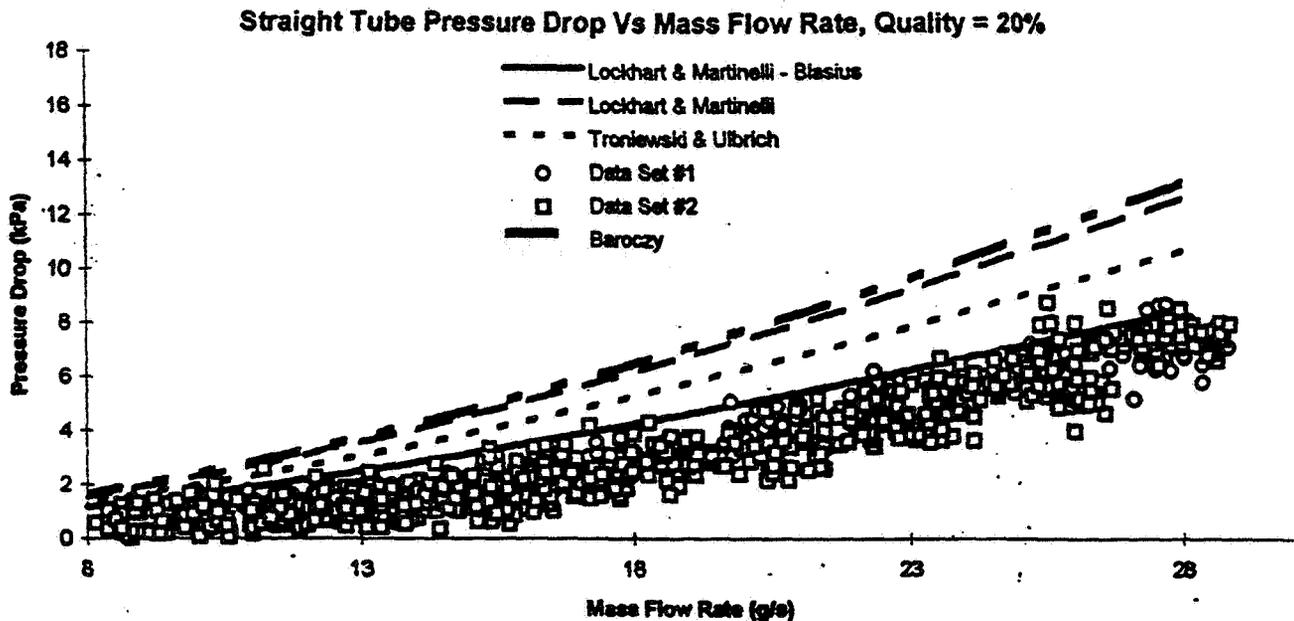


Fig. 3 Straight Tube Pressure Drop Vs. Mass Flow Rate, Quality = 20%

Figure 4 illustrates the measured pressure drop in the straight tube section for two-phase R-113 flow subjected to evaporator input levels which produced a flow quality of 30% as well as the corresponding predictive correlations. As observed in Fig. 3, each of the correlations produced a prediction of pressure drop which was greater than that observed experimentally. The most significant difference between Fig. 4 and Fig. 3 is that the predictive correlation which utilized the Blasius Solution¹ friction factor with the prediction of Lockhart and Martinelli² demonstrated a significant "jump" at the transition to turbulent flow of the R-113 liquid in the two-phase flow. The predictive correlation of Troniewski and Ulbrich³ yielded the closest agreement with the measured values

Figure 5 illustrates the measured and predicted pressure drop values for two-phase flow having a quality of 40%. While two of the correlations, that of Troniewski and Ulbrich³ and that of Lockhart and Martinelli² match both the trends and values of the measured data, this may have been simply because it is at this quality value that the fluctuation in measured mass flow rate became quite significant. At this quality value, fluctuations in the measured value of the mass flow rate were as great as $\pm 15\%$. While this fluctuation occurred in the measured value, the trends of the data displayed in Fig. 4 indicate that there was little effect on the measured pressure drop values lying within the "high end" of

the range of data. However, for the measured pressure drop values lying within the low end of the data (< 8 kPa), the uncertain fluctuation in mass flow rate appeared to be accompanied by an uncertain fluctuation in the measured pressure drop value.

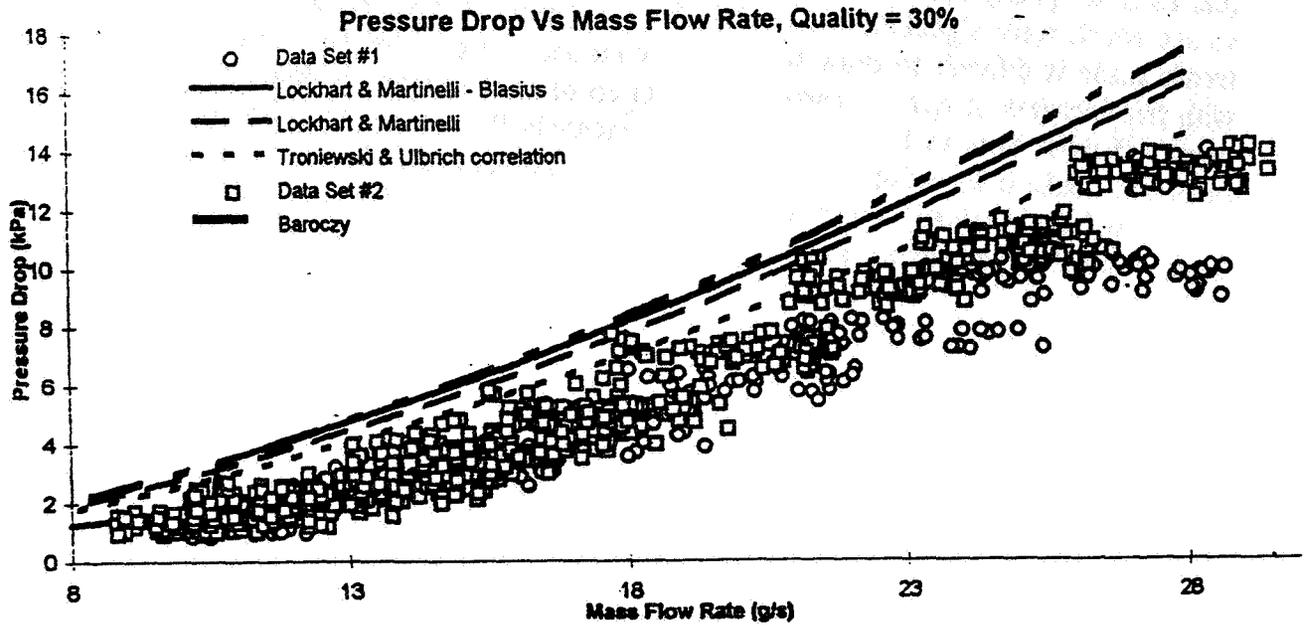


Fig. 4 Straight Tube Pressure Drop Vs. Mass Flow Rate, Quality = 30%

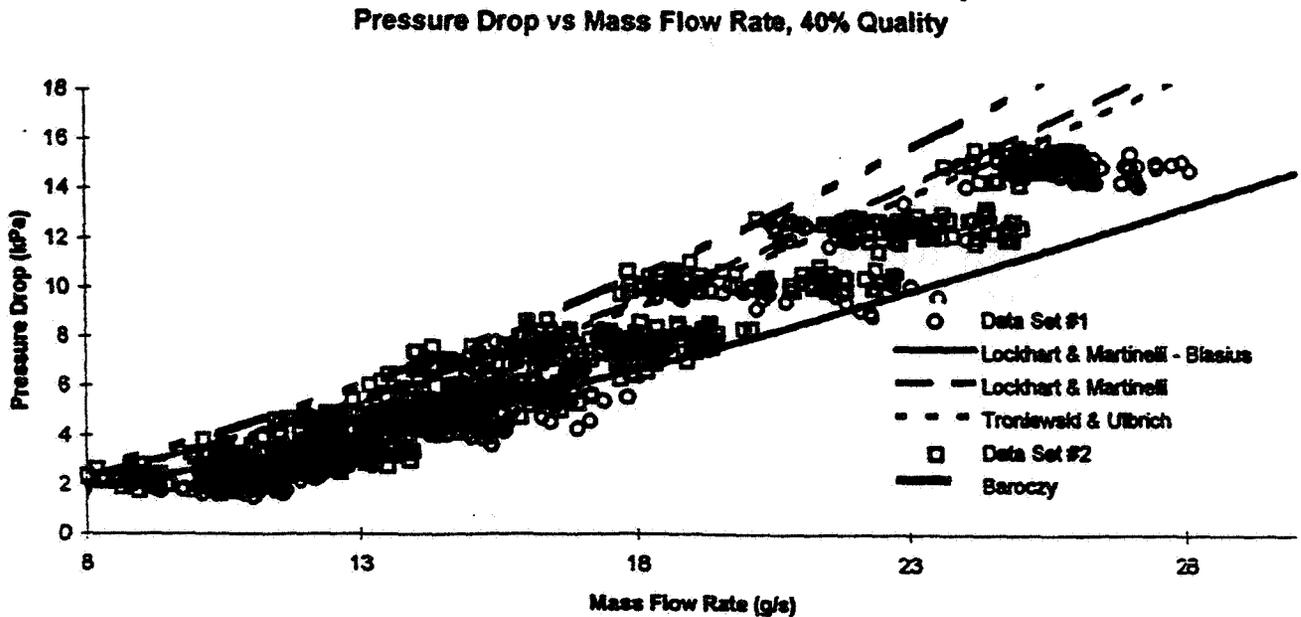


Fig. 5 Straight Tube Pressure Drop Vs. Mass Flow Rate, Quality = 40%

Figures 6 - 9 illustrate the measured and predicted pressure drop values for two-phase flows having qualities of 50%, 60%, 70%, and 80%, respectively. In general, it appears that as flow quality increased, each of the correlations began to predict pressure drop values which were significantly less than those measured experimentally. Experimental trends made it difficult to draw worthwhile conclusions from the data. Unfortunately, with this increase in quality, came a great increase in the uncertain fluctuation of the measured mass flow rate value. This large uncertain fluctuation made the process of obtaining useful information from the experiments quite difficult. Only the most general trend could be observed, which was that of the predictive correlations yielding pressure drop values less than those measured experimentally.

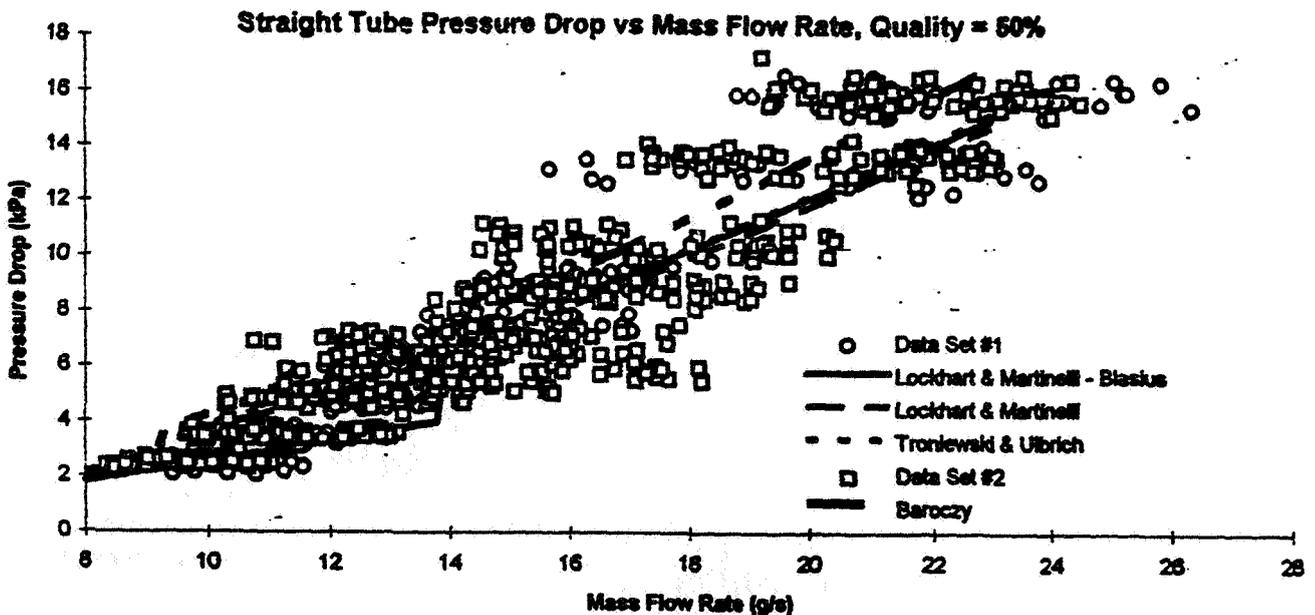


Fig. 6 Straight Tube Pressure Drop Vs. Mass Flow Rate, Quality = 50%

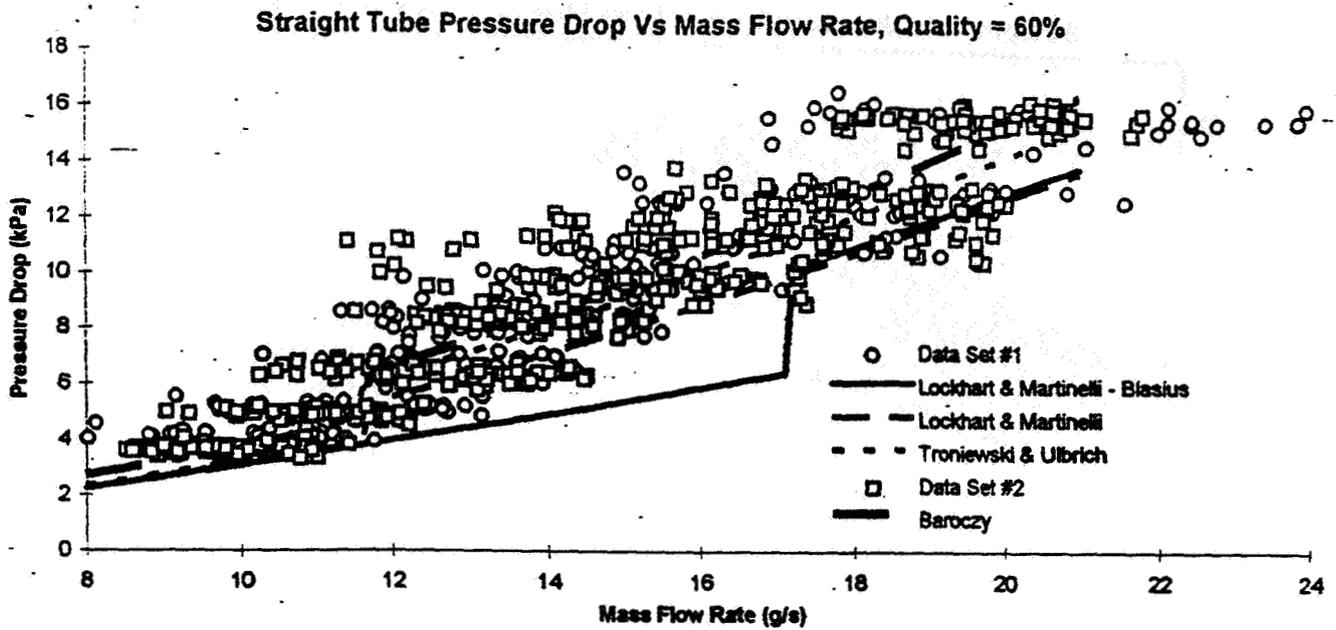


Fig. 7 Straight Tube Pressure Drop Vs. Mass Flow Rate, Quality = 60%

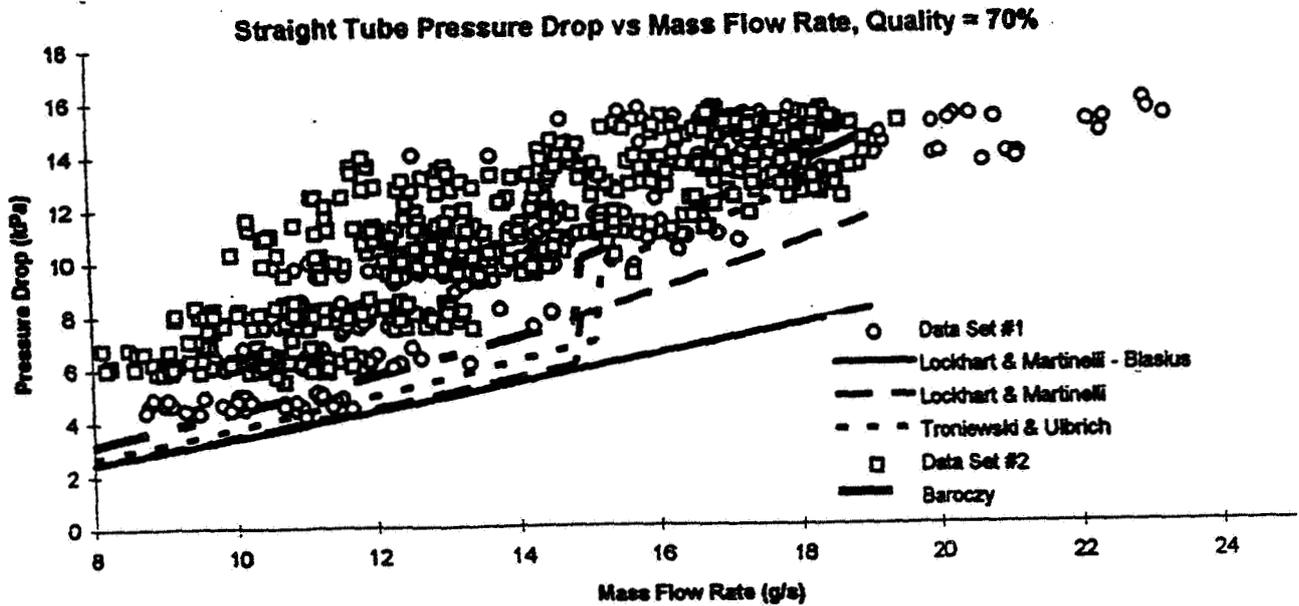


Fig. 8 Straight Tube Pressure Drop Vs. Mass Flow Rate, Quality = 70%

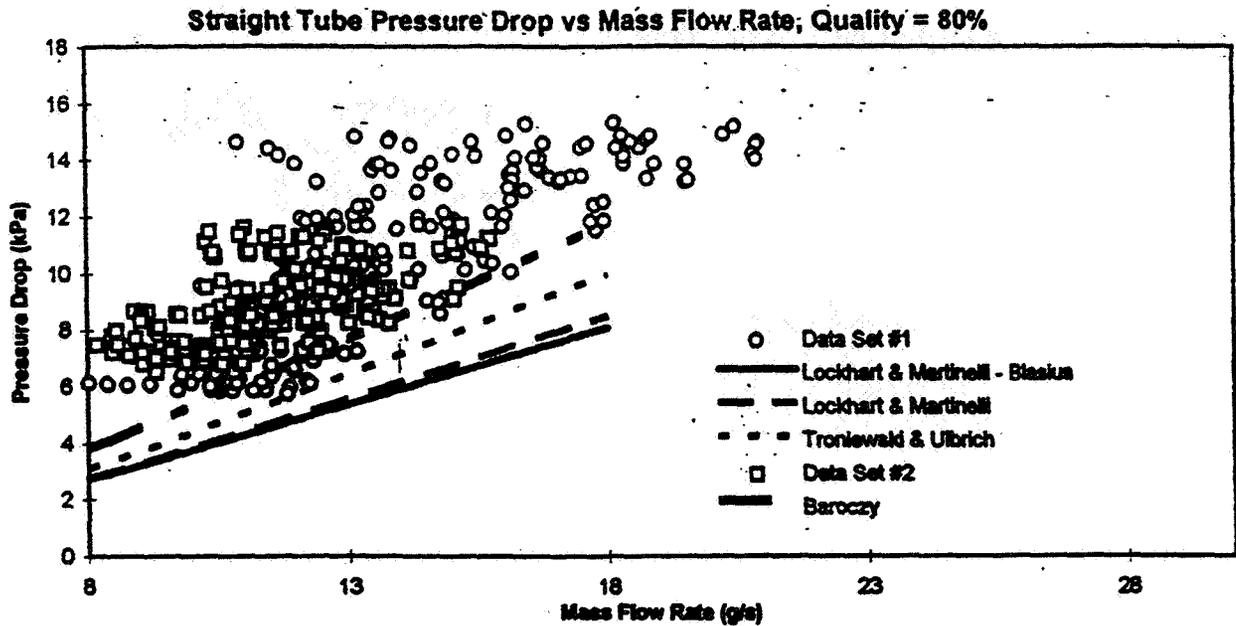


Fig. 9 Straight Tube Pressure Drop Vs. Mass Flow Rate, Quality = 80%

Investigation of Two-Phase Flow Through Piping Bends

The test facility was modified in an attempt to obtain pressure drop information for flow through 180° bends, as illustrated in Fig. 10. Unfortunately, this attempt was not successful. The lack of success is primarily due to two different system design characteristics. First, it was concluded that fluctuations mass flow (and therefore quality) due to percolation in the evaporator section produced random changes in system pressure which were greater in magnitude than those measured across the tubing bend. Second, the configuration of pressure transducers prescribed for the system did not allow for accurately and repeatably obtaining pressure drop data across the tubing bend.

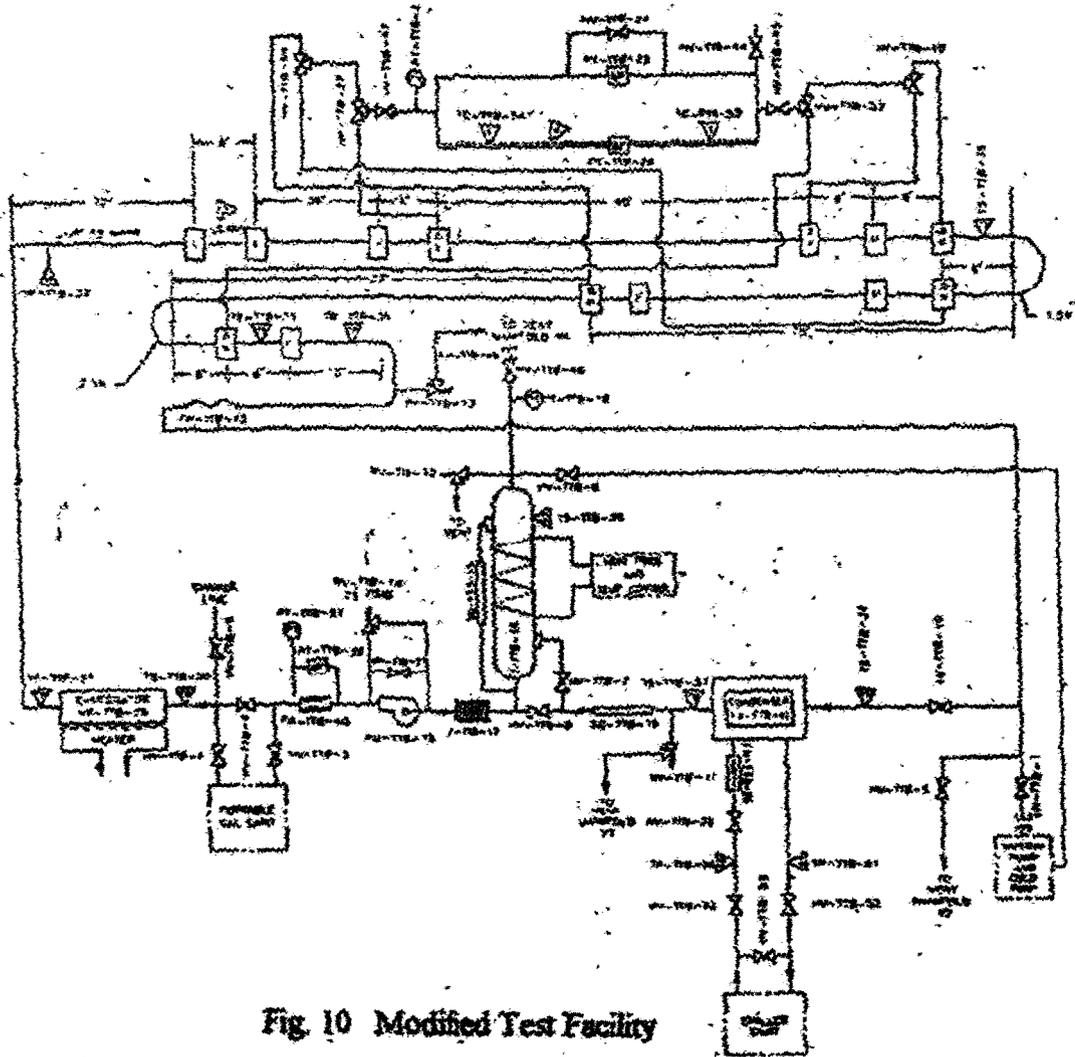


Fig. 10 Modified Test Facility

Conclusions and Recommendations for Future Research Efforts

Data has been obtained for two-phase flow of R-113 through $\frac{1}{2}$ " diameter straight tubing and tubing bends. The pressure drop data obtained for flow through straight tubing sections demonstrated trends which agreed with those of the predictive models. Each of the three models utilized demonstrated reasonable agreement with the data within a limited range of quality, though none of the three models agreed with the data throughout the entire range of qualities observed (0% - 80%).

The measurements of pressure drop through the tubing bends illustrated unacceptable levels of uncertainty. This was due to uncertain random fluctuations occurring in the mass flow rate in the test facility. Based on the observation of high speed video films of the flow, this fluctuation in the mass flow rate was due to a percolating pressure gradient which was produced in the vertical evaporator section. This fluctuation adversely affected the measurement of pressure drop in the tubing bend due to the fact that the magnitude of the pressure drop through the bend was less than the magnitude of the random pressure fluctuation induced by the percolation.

Future efforts should initially be directed at eliminating the fluctuation in the mass flow rate associated with the two-phase test facility. Replacing the vertical evaporator section with a horizontal section having significantly greater effective tubing length and lower heat

flux values should eliminate the occurrence of percolation. Additionally, replacing the current flat plate orifice flow meter with a rotor-type flow meter may reduce the noise level associated with the signal delivered to the data acquisition system.

Beyond reconfiguration efforts for the two-phase test facility, several investigations should be undertaken. Certainly, the first of these should be to complete a comprehensive investigation of the pressure drop characteristics of two-phase flows through the 180° bends of the modified facility. Investigations utilizing various refrigerants, mass flux values, and tubing diameters should be undertaken. Additionally, investigations of the pressure drops measured through quick-disconnects and other fittings, as well as tubing expansions, contractions, and manifolds should follow. Each of these experimental investigations should be coupled with rigorous analysis of the observed phenomena.

References

1. Blasius, H., "Grenzschichten in Flüssigkeiten mit kleiner Reibung," *Z. Math. Phys.*, Vol. 56, 1908, p. 1; also NACA TM 1256.
2. Lockhart, R. C., and Martinelli, R. W., "Proposed Correlation of Data for Isothermal Two-Phase Two Component Flow in Pipes," *Chem. Eng. Prog.*, Vol. 45, No. 1, pp. 39 - 48, 1949.
3. Troniewski, L., and Ulbrich, R., "Two-Phase Gas Liquid Flow in Rectangular Channels," *Chem. Engr. Sci.*, Vol. 39, No. 4, pp. 751-765.
4. Baroczy, C. J., "A Systematic Correlation for Two-Phase Pressure Drop," *Chem. Engr. Prog. Ser.*, Vol 64, pp. 232-249.
5. Chisolm, D., "Brief Communications - Two-Phase Flow in Bends," *Int. J. Multiphase Flow*, Vol. 6, August, 1980, pp. 363-367.

**A VELOCITY DISTRIBUTION MODEL
FOR STEADY STATE HEAT TRANSFER**

**Final Report
NASA/ASEE Summer Faculty Fellowship Program—1995
Johnson Space Center**

Prepared By:	Eric B. Hall, Ph.D.
Academic Rank:	Assistant Professor
College and Department:	Southern Methodist University Dept. of Electrical Engineering Dallas, Texas 75275
NASA/JSC	
Directorate:	Space Biomedical Research Institute
Division:	Medical Sciences Division
Branch:	Space and Life Sciences
JSC Colleague:	Alan H. Feiveson, Ph.D.
Date Submitted:	August 4, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

Consider a box that is filled with an ideal gas and that is aligned along Cartesian coordinates (x, y, z) having unit length in the “ y ” direction and unspecified length in the “ x ” and “ z ” directions. Heat is applied uniformly over the “hot” end of the box ($y = 1$) and is removed uniformly over the “cold” end ($y = 0$) at a constant rate such that the ends of the box are maintained at temperatures T_0 at $y = 0$ and T_1 at $y = 1$. Let U , V , and W denote the respective velocity components of a molecule inside the box selected at some random time and at some location (x, y, z) . If $T_0 = T_1$, then U , V , and W are mutually independent and Gaussian, each with mean zero and variance RT_0 , where R is the gas constant. When $T_0 \neq T_1$ the velocity components are *not* independent and are *not* Gaussian. Our objective is to characterize the joint distribution of the velocity components U , V , and W as a function of y , and, in particular, to characterize the distribution of V given y . It is hoped that this research will lead to an increased physical understanding of the nature of turbulence.

THE SIMULATION

Consider a box that is filled with an ideal gas and that is aligned along Cartesian coordinates (x, y, z) having until length in the “ y ” direction and unspecified length in the “ x ” and “ z ” directions. Heat is applied uniformly over the “hot” end of the box ($y = 1$) and is removed uniformly over the “cold” end ($y = 0$) at a constant rate such that the ends of the box are maintained at temperatures T_0 at $y = 0$ and T_1 at $y = 1$. Let U , V , and W denote the respective velocity components of a molecule inside the box selected at some random time and at some location (x, y, z) . (See Figure 1.) Our concern in this project is to characterize these velocity components. In this regard, we will first consider the problem of simulating the molecular motion inside the box so as to obtain simulated velocities. This simulation will be based on the direct simulation Monte Carlo code presented in [Bird, 1994].

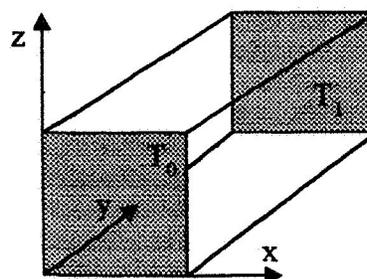


Figure 1

For the simulation, we will consider (as indicated by Figure 2) a thin slice of the box that effectively ignores the z spatial component. The z component is not entirely neglected however since the molecule velocities are considered three dimensional even though the molecule positions are taken to be only two dimensional. This two dimensional slice is divided into cells and subcells as indicated by Figure 3. As Bird describes the problem in [Bird, 1994, p. 215], subcells were introduced because it “was feared that coincidental collisions between molecules at opposite sides of the cells, coupled with the immediate migration of these molecules to the next cell, might cause false disturbances to propagate with speed equal to the ratio of the cell size to the time step.”

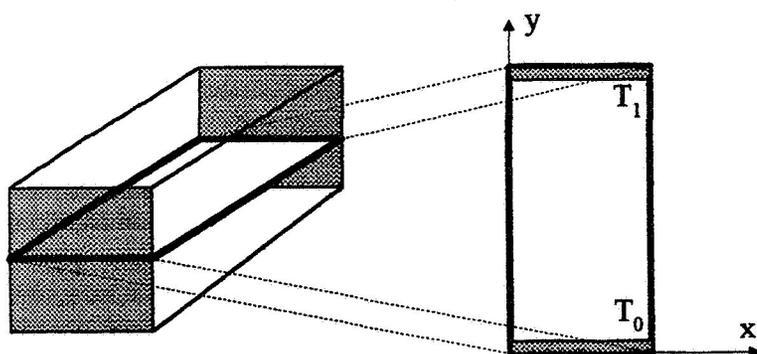


Figure 2

by the program. The line numbers mentioned below refer to the copy of this file that is reproduced at the end of this report.

equal to the ratio of the cell size to the time step.”

In order to explain the operation of the simulation program, we will consider the input file `dsmc2.dat` that contains the data used to produce a particular simulation run. Any line in the data file beginning with ‘#’ is treated as a comment and is ignored

The first item in the file (lines 6-9) determines whether or not an output file `dsmc2.out` is produced. This file corresponds to the only output provided by the original program `dsmc2.for` from [Bird, 1994].

When the program begins, it first asks whether a new calculation is desired or an old calculation should be restarted. This latter option is only possible when line 19 of the data file has been set to 0 in an earlier run in order to produce a restart file. The restart file `dsmc2.res` contains a snapshot of the current state of the simulation and can be used to restart the program at that point should any sort of interruption occur. (Note that the size of this restart file can be as large as 20 Meg.) For long simulations, it is suggested that the values in lines 19-20 be chosen so that a restart file is produced approximately once each hour. Producing the file more frequently may greatly reduce the speed of the simulation.

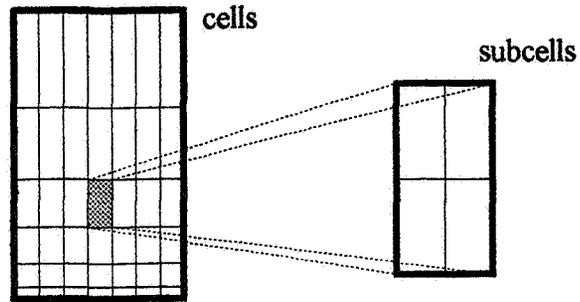


Figure 3

A second output option is the production of a log file `dsmc2.log` that contains velocity and temperature information for cells along a particular y location. This file is produced when line 26 of the data file is set to zero. The value for y is specified by line 150 of the input file. In addition, this log file contains information regarding the number of molecules per cell that may be useful is choosing a proper value of $CWRY$, which is considered below. A sample of this output file is given near the end of this report. Note that the averages given by this option are not cumulative but are instead based only on the number of molecules occupying cells along the given y location at the time the file is updated.

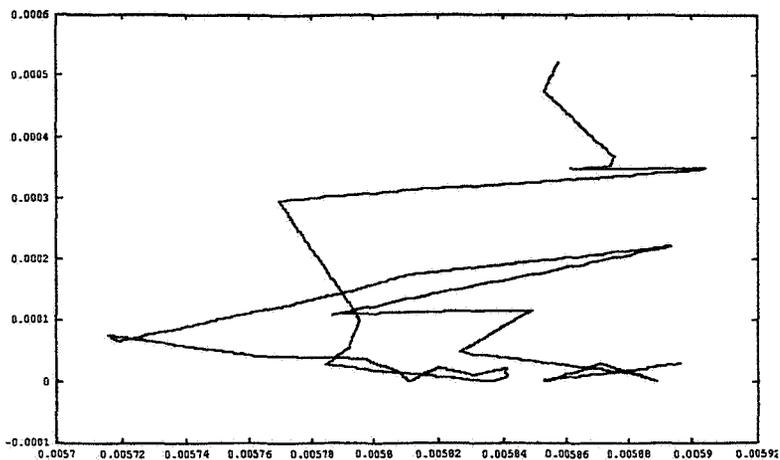
The next output option determines whether or not a chart of velocity moments is produced during the simulation. When this option is selected by setting line 31 of the data file to zero, a chart as given later in this report is periodically written to the screen and to the file `dsmc2.cht`. The statistics in this chart are cumulative over the run of the simulation. The third column indicates the total number of molecules that have been used to obtain the sample moments found in the final three columns. The row numbers that are used to produce the chart are determined by lines 163-182 of the data file. Note that it is possible to produce this data for every row in the box simply by setting line 167 in the data file equal to the total number of rows in the box.

A third output option makes it possible to track the position of a molecule initially located along some specified row. A file called `dsmc2.mol` containing the (x, y) coordinates of a molecule in the specified row is produced when line 36 of the data file is set to zero. The initial molecule row number is specified by line 154 of the data file. Since the molecules generally move only slightly between successive time steps, the step size provided

by line 161 makes it possible to more easily obtain a picture of the molecule's motion over a longer period of time.. A graph that shows how the data file `dsmc2.mol` can be used to study the molecular motion produced by the simulation is given below.

The cell widths in the y direction are not uniform, but instead are chosen to be smaller at the cold end of the box than at the hot end of the box. The ratio $CWRY$ of the cell width at the hot end of the box to the cell width at the cold end is given in line 40 of the data file. If

$S = CWRY^{1/199}$, then the width of the cell in Row i is given by $WS^{i-1}(1-S)/(1-S^{200})$ where W is the total width of the box in the y direction. Note that here we have assumed that the total number of rows in the y direction is 200. This value can be changed if the code is recompiled. Generally, the ratio $CWRY$ should be the square root of the ratio of the temperatures at the two ends of the boxes. A graph showing the number of molecules in each row for a particular simulation is given near the end of this report.



The initial temperature of the gas is determined by $FTMP$, which we have chosen to be the geometric mean of the two surface temperatures. This value is given in line 44 of the data file. The initial number density FND is in molecules per square meter. This value is specified in line 48 of the data file.

In the DSMC method, a single simulated molecule represents $FNUM$ actual molecules. This value is given in line 52 of the data file and must be chosen so that the total number of simulated molecules is not greater than the maximum allowed number of molecules, which in this case is 40000. (This value can be changed if the code is recompiled.) The actual number of simulated molecules used by the program is given by $(FND/FNUM)$ multiplied by the area of the box.

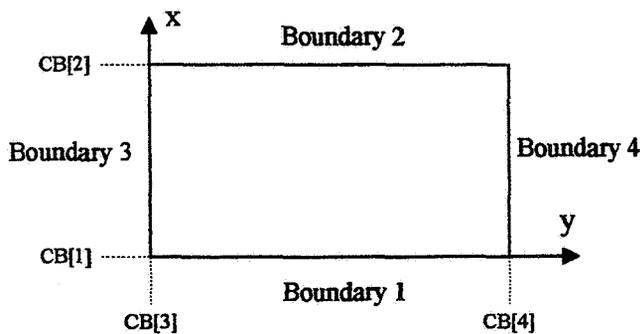


Figure 4

The value DTM in line 56 of the data file is the time step of the simulation, and it should be chosen so that the molecular motion and the collisions are uncoupled. In particular, it should be much less than the local mean collision time. The DSMC

method becomes increasingly more accurate as the cell size and the time step tend to zero. (See [Bird, 1994, p. 215].)

The values $CB[i]$ for $i = 1, 2, 3,$ and $4,$ are specified in lines 69-81 of the data file and

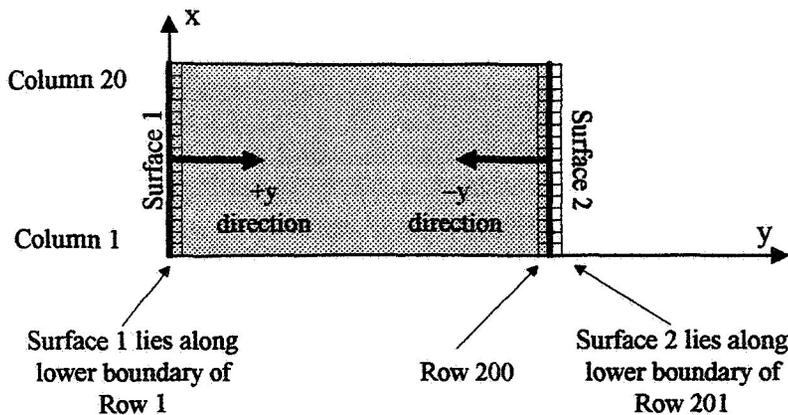


Figure 5

determine the dimensions of the box as indicated by Figure 4. Two planar surfaces (Surface 1 and Surface 2) at possibly different temperatures may be placed anywhere within the box. The reflection at these surfaces is diffuse, which means that the return velocity of a colliding molecule is determined by the temperature of the

wall instead of by the incident velocity. At the walls (or boundaries) of the box, the reflection is specular, which means that the appropriate velocity component before the collision is simply reversed to obtain the velocity component after the collision.

Surface 1 is specified by the parameters in lines 92-110 of the data file. The value $ISURF[1]$ specifies the direction of a line normal to the surface and pointing toward the gas. For specular reflection, it is this velocity component that is reversed upon reflection. (The velocity components parallel to the surface are not changed.) For example, in Figure 5, such a normal is in the positive y direction, which corresponds to a value of 1 for $ISURF[1]$. The negative y direction corresponds to a value of 2, the positive x direction corresponds to a value of 3, and the negative x direction corresponds to a value of 4. The values of $LIM[1][i]$ specify the location of Surface 1 as indicated by Figure 5 and the descriptions in lines 96-106 of the data file. The value $TSURF[1]$ is the temperature of Surface 1. Surface 2 is similarly specified by the parameters in lines 112-130.

The values in lines 132-146 determine the rate of output and the duration of the simulation. Their use is illustrated by the flow chart near the end of this report. Note that the sampling does not begin until NPS output cycles have occurred. The moment chart is not produced until this same point is reached.

FUTURE DIRECTIONS

The ultimate goal of this work is to characterize the conditional distribution of the y component of the velocity given a particular location y . Since this problem appears to be

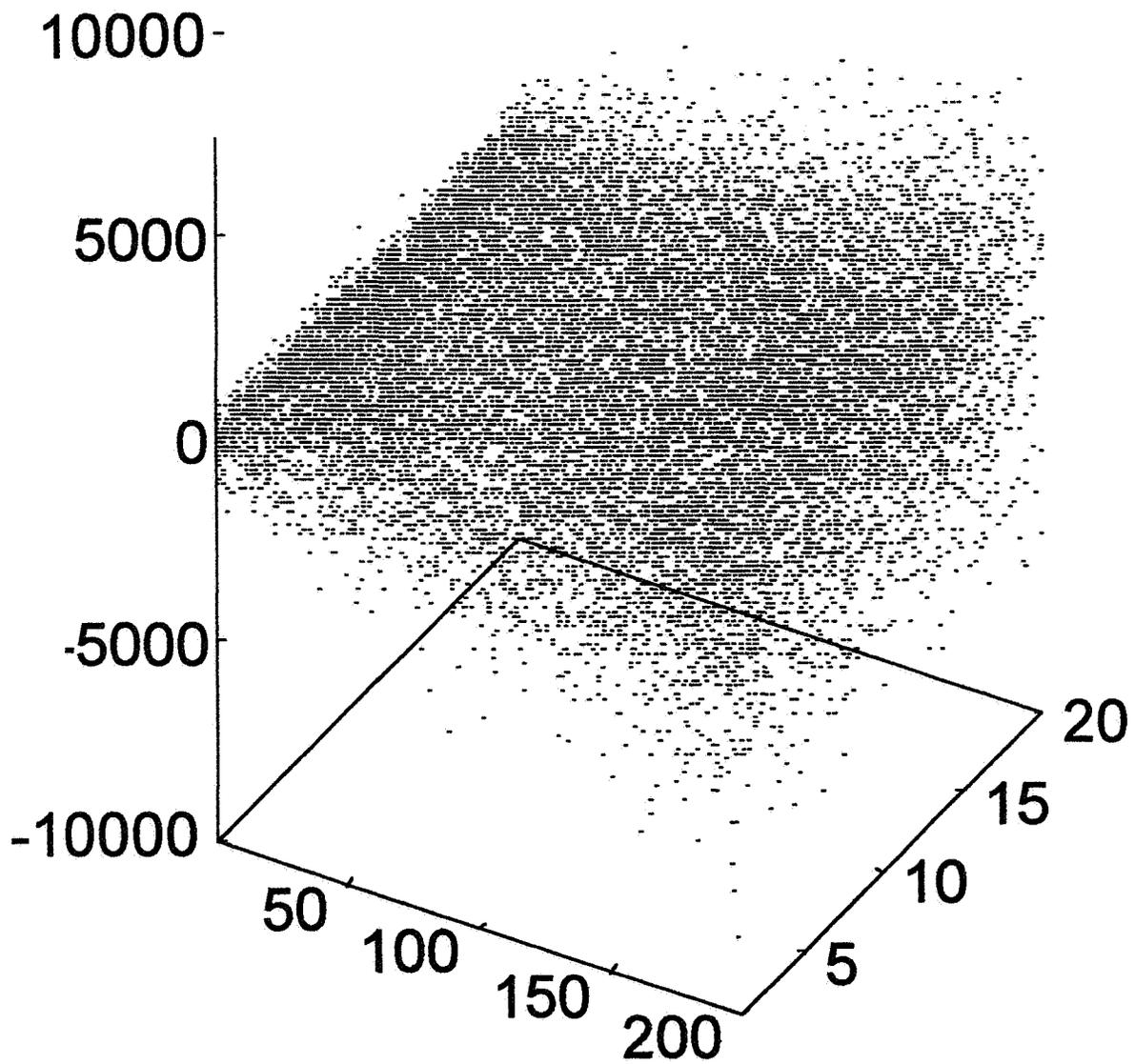
analytically intractable, any practical investigation of the distribution will involve data obtained either empirically or via a simulation such as the one that we have considered here.

One immediate item for future consideration involves the effect of the random number generator on the simulation results. We, as did Bird, have used the algorithm by Knuth that is found in [Press *et al.*, 1992, p. 283]. In [Ferrenberg *et al.*, 1992], it was shown that subtle numerical errors can occur in Monte Carlo simulations when many supposedly ‘good’ random number generators are used. Since our simulation requires a *large* number of random numbers, it would be helpful to repeat and compare these simulations with a variety of different random number algorithms.

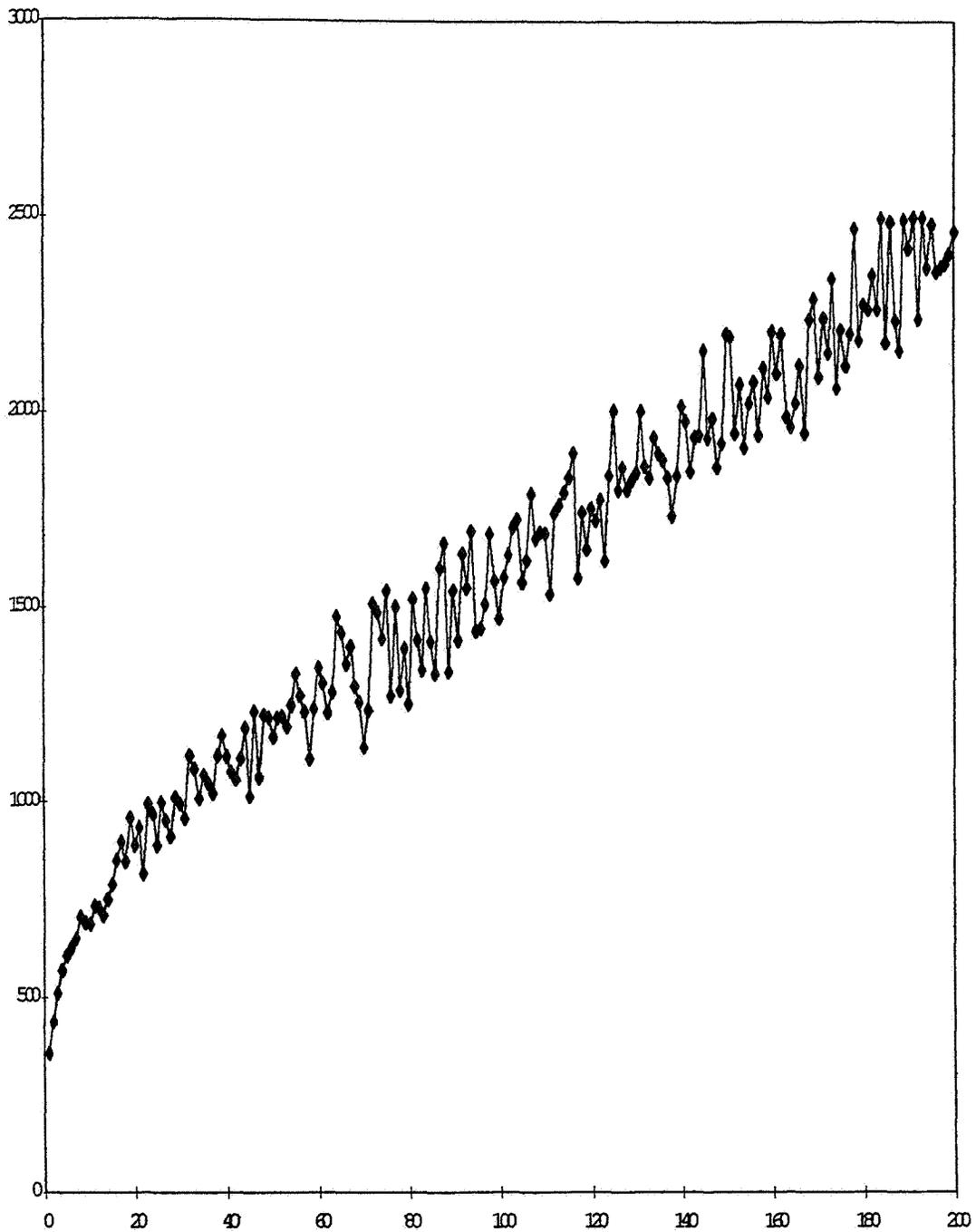
A second item for future consideration involves an investigation of the boundary conditions at the two surfaces. The current code assumes that the distribution of the velocity component that is normal to the surface has the form of a zero mean Gaussian distribution taken to the right of the origin and renormalized. This is the distribution that would correspond to molecules crossing the surface from some external stationary equilibrium gas. (See [Bird, 1994, p. 259].) An alternate model that provides a better fit with the simulated data is a distribution obtained by taking the absolute value of the difference of two independent gamma random variables.

REFERENCES

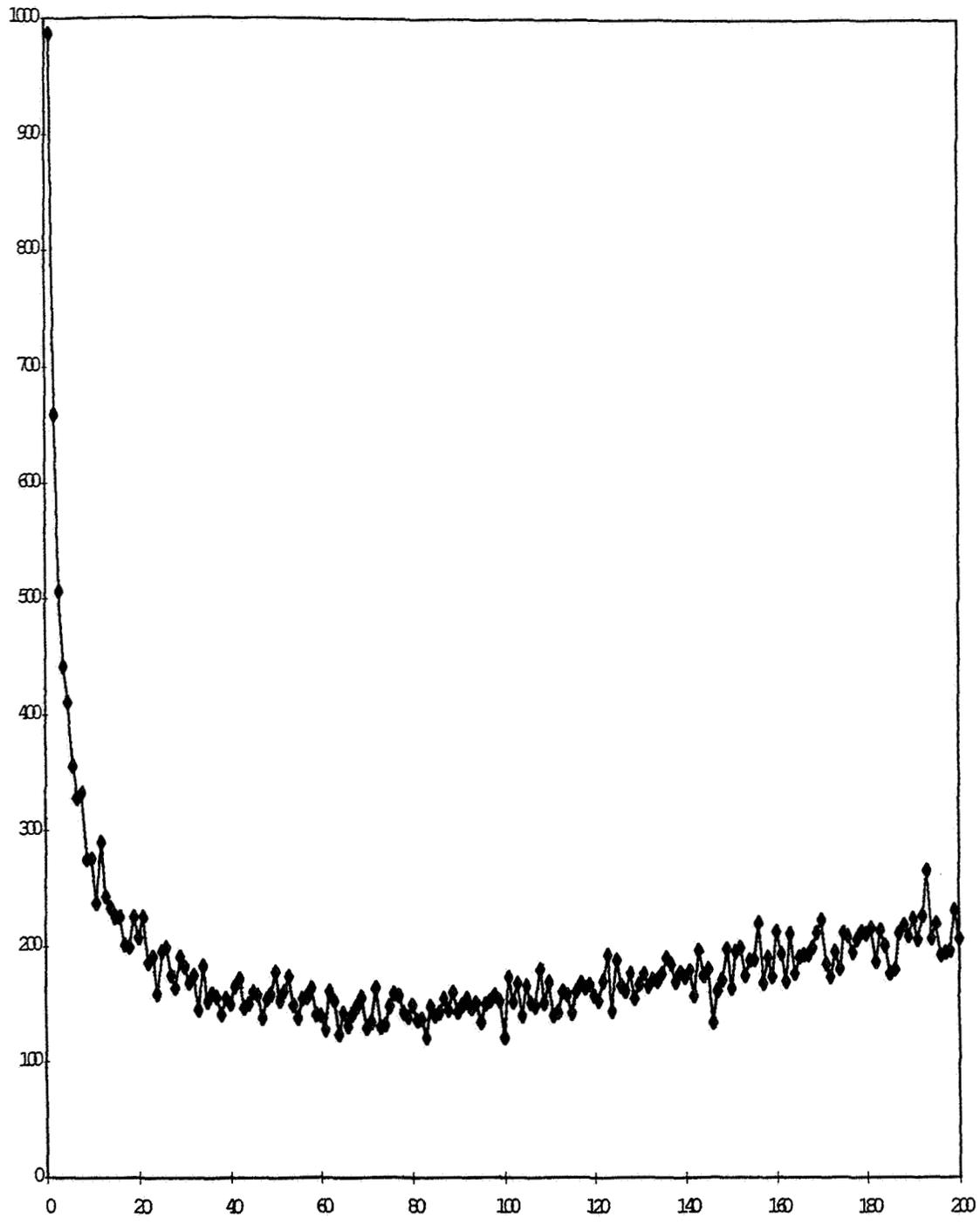
- Bird, G. A., *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, Clarendon Press, Oxford, 1994.
- Chapman, S., and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, Third edition, Cambridge University Press, Cambridge, 1970.
- Ferrenberg, A. M., D. P. Landau, and Y. Joanna Wong, “Monte Carlo Simulations: Hidden Errors from ‘Good’ Random Number Generators,” *Physical Review Letters*, Vol. 69, No. 23, Dec. 7, 1992, pp. 3382–3384.
- Meyer, R. E., *Introduction to Mathematical Fluid Dynamics*, Dover Publications, New York, 1971.
- Press, W., S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Second edition, Cambridge University Press, 1992.
- Prigogine, I., “Evolution Criteria, Variational Properties, and Fluctuations,” *Nonequilibrium Thermodynamics: Variational Techniques and Stability*, University of Chicago Press, Chicago, 1966, pp. 3-16.



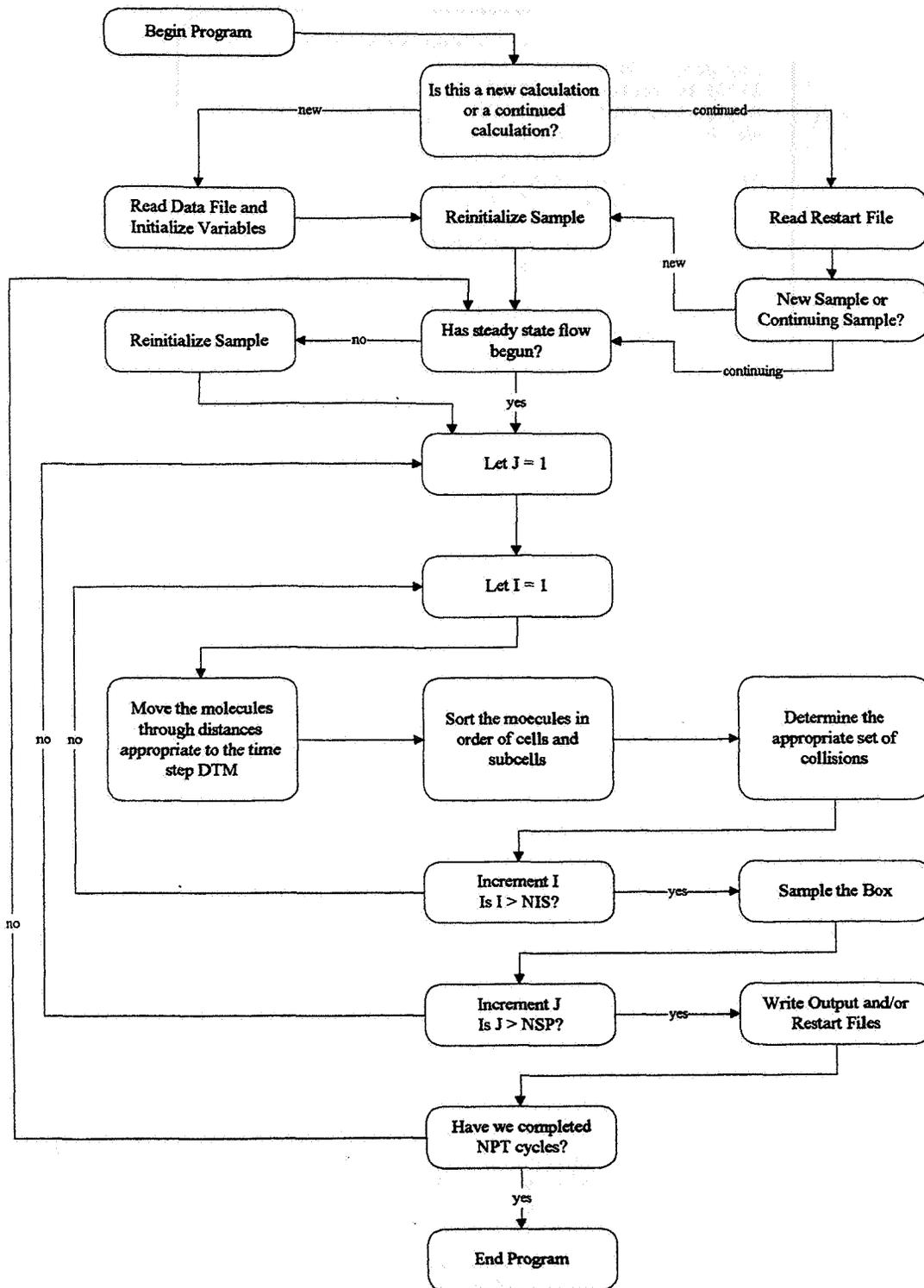
Scatter plot of y velocity component in meters/second for approximately 16000 molecules selected at random from the box after approximately 268 million collisions have occurred. The temperature at row 1 is 300 and at row 200 is 30000.



A plot of the root mean square value of the y velocity component for each row in the box. Again, the values were obtained for a box that had experienced approximately 268 million simulated collisions. The temperature at row 1 is 300 and at row 200 is 30000.



A plot of the number of molecules present in each row of a box that has experienced approximately 268 million simulated collisions. The cell width at row 200 is ten times the cell width at row 200.



```

*****
Log Update 94
37538 Molecules and 432153 Collisions
Average molecules per cell = 9.4
Maximum molecules per cell = 37
Minimum molecules per cell = 0
Percentage of nonzero cells = 99%
Data for cells along line y = 0.005840
Total number of molecules = 176
Average Temperature of Cells = 2932.75
Average U velocity component = 0.04133742 km/s
Average V velocity component = 0.002432757 km/s
Average W velocity component = -0.002770768 km/s
Average Squared U vel compt = 0.5711446
Average Squared V vel compt = 0.719397
Average Squared W vel compt = 0.6397187
Average U^3 velocity compt = 0.101235
Average V^3 velocity compt = 0.01372038
Average W^3 velocity compt = -0.04497306
Average U^4 velocity compt = 0.9689213
Average V^4 velocity compt = 1.44099
Average W^4 velocity compt = 1.14349
Average X Component of Mol = -0.0002265342
Average Y Component of Mol = 0.005869242

```

Sample of dsmc2.log

```

Molecules = 37538
Collisions = 145790602
Start Time = 1.16e-06
End Time = 0.00055648
DTM = 2e-09
FNUM = 1.3e+14
CWRV = 10
FTMP = 3000
FND = 0
Surf 1 Temp = 300
Surf 2 Temp = 30000

```

Row	Y Midpoint	Mol Count	2nd Moment	3rd Moment	4th Moment
20	0.00055534	2874478	0.7194	-0.2035	1.7523
40	0.00127118	2260561	1.1495	-0.3343	4.3424
50	0.00169623	2158229	1.3561	-0.3945	6.0186
60	0.00217341	2103919	1.5588	-0.4584	7.9019
80	0.00331058	2086605	1.9812	-0.5631	12.6606
100	0.00474384	2146617	2.4251	-0.7288	18.9076
120	0.00655029	2246682	2.9223	-0.8890	27.4045
140	0.00882712	2378198	3.4810	-1.0843	38.6212
160	0.01169679	2527745	4.1265	-1.3110	53.6066
180	0.01531368	2695610	4.8761	-1.5899	73.5440

Sample of dsmc2.cht

```

1 | #-----
2 | # DATA FILE FOR DSMC2
3 | # Any line beginning with '#' is ignored.
4 | # Data should be placed on a single line following each description.
5 | #-----
6 | # 1: Do not produce an output file
7 | # 0: Produce output files
8 | #-----
9 | 1
10 | #-----
11 | # 1: Do not produce a restart file
12 | # 0: Produce a restart file at every Nth output file
13 | # The value of N should follow the flag value below.
14 | # For example, if the next two data lines were
15 | #     0
16 | #     5
17 | # then a restart file would be produced after each group of 5 outputs.
18 | #-----
19 | 0
20 | 100
21 | #-----
22 | # 1: Do not produce a log file
23 | # 0: Produce a log file whenever an output file is produced
24 | # See Y Value below
25 | #-----
26 | 1
27 | #-----
28 | # 1: Do not produce a moment chart
29 | # 0: Produce a moment chart whenever an output file is produced
30 | #-----
31 | 0
32 | #-----
33 | # 1: Do not produce a molecule track file
34 | # 0: Track a molecule in row specified below
35 | #-----
36 | 0
37 | #-----
38 | # CWRV: Ratio of cell width at outer to that at inner y boundary
39 | #-----
40 | 10
41 | #-----
42 | # FTMP: The stream temperature
43 | #-----
44 | 3000
45 | #-----
46 | # FND: The initial number density
47 | #-----
48 | 1.22e22
49 | #-----
50 | # FNUM: The number of real mols represented by each simulated mol
51 | #-----
52 | 0.13e15
53 | #-----
54 | # DTM: The time step over which mol motion and collisions are uncoupled
55 | #-----
56 | 2e-9
57 | #-----
58 | #
59 | #
60 | #
61 | #
62 | #

```

	Boundary 2	
high x		
Boundary		Boundary
3		4

```

63 | #           |           |
64 | #           | low x -----|
65 | #           |           | low y   Boundary 1   high y
66 | #-----|
67 | # CB[1]: The x coordinate of boundary 1 (low x)
68 | #-----|
69 | -0.01
70 | #-----|
71 | # CB[2]: The x coordinate of boundary 2 (high x)
72 | #-----|
73 | 0.01
74 | #-----|
75 | # CB[3]: The y coordinate of boundary 3 (low y)
76 | #-----|
77 | 0
78 | #-----|
79 | # CB[4]: The y coordinate of boundary 4 (high y)
80 | #-----|
81 | 0.02
82 | #-----|
83 | #
84 | #           | Col 20 -----|
85 | #           |           |
86 | #           | Surface |----> pos y       neg y <----| Surface
87 | #           | 1      |      dir           dir           | 2
88 | #           |           |
89 | #           | Col 1 -----|
90 | #           |           | Row 1           Row 200
91 | #-----|
92 | # ISURF[1]: Direction of surface 1 normal is in the positive y direction
93 | #-----|
94 | 1
95 | #-----|
96 | # LIMS[1][1]: Surface 1 lies along lower boundary of this row
97 | #-----|
98 | 1
99 | #-----|
100 | # LIMS[1][2]: Surface 1 begins at column 1
101 | #-----|
102 | 1
103 | #-----|
104 | # LIMS[1][3]: and continues until column 20
105 | #-----|
106 | 20
107 | #-----|
108 | # TSURF[1]: Temperature of surface 1
109 | #-----|
110 | 300
111 | #-----|
112 | # ISURF[2]: Direction of surface 2 normal is in the negative y direction
113 | #-----|
114 | 2
115 | #-----|
116 | # LIMS[2][1]: Surface 2 lies along lower boundary of this row
117 | #-----|
118 | 201
119 | #-----|
120 | # LIMS[2][2]: Surface 2 begins at column 1
121 | #-----|
122 | 1
123 | #-----|
124 | # LIMS[2][3]: and continues until column 20

```

```

125 | #-----
126 | 20
127 | #-----
128 | # TSURF[2]: Temperature of surface 2
129 | #-----
130 | 30000
131 | #-----
132 | # NIS: The number of DTM time steps between samplings
133 | #-----
134 | 2
135 | #-----
136 | # NSP: The number of samples between prints
137 | #-----
138 | 10
139 | #-----
140 | # NPS: The number of prints until assumed start of steady flow
141 | #-----
142 | 30
143 | #-----
144 | # NPT: The total number of prints
145 | #-----
146 | 50000
147 | #-----
148 | # Y value for which output will be collected in the LOG FILE
149 | #-----
150 | 0.00584
151 | #-----
152 | # Row Number Containing Molecule to Track
153 | #-----
154 | 100
155 | #-----
156 | # Step size for molecule position output to file
157 | # i.e. if next data line is 10 then position information will be
158 | # written to the file every 10 time steps
159 | # Use 1 to output a position after every time step
160 | #-----
161 | 10
162 | #-----
163 | # Number of Rows for which Output is Desired.
164 | # If this value equals or exceeds the total number of rows then data
165 | # will be output for all rows.
166 | #-----
167 | 10
168 | #-----
169 | # Row Numbers to Output.
170 | # Required if previous data line is greater than zero and less than the
171 | # total number of rows.
172 | #-----
173 | 20
174 | 40
175 | 50
176 | 60
177 | 80
178 | 100
179 | 120
180 | 140
181 | 160
182 | 180

```

TRABECULAR BONE MECHANICAL PROPERTIES AND FRACTAL DIMENSION

**Final Report
NASA/ASEE Summer Faculty Fellowship Program--1995
Johnson Space Center**

Prepared By: Harry A. Hogan, Ph.D.

Academic Rank: Associate Professor

University & Department: Texas A&M University
Department of Mechanical Engineering
College Station, Texas 77843-3123

NASA/JSC

Directorate: Space and Life Sciences

Division: Medical Sciences

Branch: Space Biomedical Research Institute

JSC Colleague: Linda C. Shackelford, M.D.

Date Submitted: August 7, 1995

Contract Number: NGT-44-001-800

ABSTRACT

Countermeasures for reducing bone loss and muscle atrophy due to extended exposure to the microgravity environment of space are continuing to be developed and improved. An important component of this effort is finite element modeling of the lower extremity and spinal column. These models will permit analysis and evaluation specific to each individual and thereby provide more efficient and effective exercise protocols. In-flight countermeasures and post-flight rehabilitation can then be customized and targeted on a case-by-case basis. Recent Summer Faculty Fellowship participants have focused upon finite element mesh generation, muscle force estimation, and fractal calculations of trabecular bone microstructure. Methods have been developed for generating the three-dimensional geometry of the femur from serial section magnetic resonance images (MRI). The use of MRI as an imaging modality avoids excessive exposure to radiation associated with X-ray based methods. These images can also detect trabecular bone microstructure and architecture. The goal of the current research is to determine the degree to which the fractal dimension of trabecular architecture can be used to predict the mechanical properties of trabecular bone tissue. The elastic modulus and the ultimate strength (or strain) can then be estimated from non-invasive, non-radiating imaging and incorporated into the finite element models to more accurately represent the bone tissue of each individual of interest. Trabecular bone specimens from the proximal tibia are being studied in this first phase of the work. Detailed protocols and procedures have been developed for carrying test specimens through all of the steps of a multi-faceted test program. The test program begins with MRI and X-ray imaging of the whole bones before excising a smaller workpiece from the proximal tibia region. High resolution MRI scans are then made and the piece further cut into slabs (roughly 1 cm thick). The slabs are X-rayed again and also scanned using dual-energy X-ray absorptiometry (DEXA). Cube specimens are then cut from the slabs and tested mechanically in compression. Correlations between mechanical properties and fractal dimension will then be examined to assess and quantify the predictive capability of the fractal calculations.

INTRODUCTION

As plans and progress continue toward establishing a permanent space station, and ultimately resuming spaceflight beyond the earth's orbit, coping with the physical and psychological demands placed upon the astronauts continues to pose a significant challenge. The musculoskeletal system in particular can undergo dramatic changes from extended exposure to the weightless environment associated with long duration spaceflight. In addressing these issues, efforts centered at the Space Biomedical Research Institute of the Johnson Space Center are continuing to focus on developing improved techniques for monitoring musculoskeletal condition and applying effective countermeasures to minimize losses and maintain adequate function. An important component of this broad-based effort is aimed specifically at developing computer models that can be customized to individuals on a case-by-case basis. Much of the initial work deals primarily with the lower limb because of its prominence in overall skeletal function. Recent Summer Faculty Fellowship participants have devised methods for non-invasively determining the size and shape of the main load-bearing bones, such as the femur & tibia (Todd, 1994), and also estimating muscle forces and joint kinematics during exercise (Figueroa, 1995). The bone dimensions and geometry are derived from magnetic resonance images (MRI) since this avoids exposure to ionizing radiation as is customary with X-ray and similar methods. Finite element models of the femur can be constructed from the geometry data and the applied loads derived from physiological muscle force data. The goal is to use such models to predict the stresses and strains within the bone, and from them identify and assess regions of weakness and potential fracture risk. This detailed and quantitative insight will allow individualized evaluation of skeletal condition (pre-, post-, and in-flight) and prescription of exercise protocols for in-flight countermeasures as well as post-flight rehabilitation. The next major step in developing this capability is to determine the material properties of the bone tissue within the bones. With methods available to generate the geometry and loads, a complete and more accurate model also requires information on the material properties of the bone or bones of interest. A major challenge at this point is to devise a way to estimate these properties from MRI data, which is a process essentially unexplored to date. The goal of the current research is therefore to evaluate the usefulness of estimating mechanical properties of trabecular bone tissue from MRI data. The study is limited to trabecular bone (as opposed to cortical bone) because MRI signals do not detect cortical bone, but this is not a severe limitation since trabecular bone is much more adaptive to changes in loading and is also more susceptible to fracture risk. The specific parameter being used to quantify trabecular bone microstructure is the fractal dimension. This quantity is calculated from the MRI following the methods developed by Acharya (1995), another recent Summer Faculty Fellowship program participant. Thus, the specific aim of the current research is to correlate mechanical properties of trabecular bone tissue with fractal dimension. The properties are measured from a series of mechanical tests on *in vitro* specimens taken from bones that have first been scanned with the MRI. This will allow direct correlation of the properties and fractal dimension from the same specific sample of tissue.

PROTOCOL AND PROCEDURES

Overview

Much of the time and effort during the summer research period has been spent formulating and refining the detailed research plan. Protocol details and various test procedure options for executing the research have been evaluated. The research involves collaboration between several different laboratories and personnel so extensive coordination is essential. Identifying all of the necessary tasks and the appropriate order and procedures has been a major focus. The main steps in the overall process are:

1. Acquire human tibia specimens and prepare for "whole bone" MRI scanning.
(at University of Texas Medical School, Orthopedics Biomechanics Laboratory)
2. MRI scan proximal portion of whole tibia using clinically available resolution.
(at Baylor College of Medicine/Methodist Hospital, Medical Physics)
3. Cut section beneath tibial plateau (3-4 cm long) and prepare for "high-resolution" MRI.
(at University of Texas Medical School, Orthopedics Biomechanics Laboratory)
4. MRI scan excised piece of proximal tibia to image trabecular architecture.
(at Baylor College of Medicine /Methodist Hospital, Medical Physics)
5. Slice proximal tibia piece into slabs approximately 1 cm thick.
(at University of Texas Medical School, Orthopedics Biomechanics Laboratory)
6. Scan slabs for bone mineral density using dual energy X-ray absorptiometry (DEXA).
(at Baylor College of Medicine /Methodist Hospital, Medical Physics)
7. X-ray slabs to provide alternate images of trabecular architecture.
(at University of Texas Medical School, Orthopedics Biomechanics Laboratory)
8. Cut slabs into cubes for mechanical testing.
(at UT Med School and/or Texas A&M University)
9. Conduct mechanical tests and calculate mechanical properties of interest.
(at UT Med School and/or Texas A&M University)
10. Determine wet and dry densities and ash weights.
(at UT Med School and/or Texas A&M University)
11. Calculate fractal dimension from MRI and X-ray images.
(at State University of New York at Buffalo, Biomedical Imaging Group)

Once the experimental tests are completed and the data analyzed, correlations will be examined between mechanical properties and other parameters. The mechanical properties of interest are the elastic moduli (in all 3 orthogonal directions) and the ultimate stress, ultimate strain, and energy absorbed along the primary axis of *in vivo* loading (i.e. the superior-inferior direction). The elastic moduli are measures of material stiffness while the other quantities are measures of strength or "ductility". The main microstructural parameter is the fractal dimension of the trabecular architecture. The fractal dimension is a novel quantity not commonly used for such purposes but recent studies have shown it to be a unique measure in distinguishing between normal and osteoporotic bone (Ruttimann *et al.*, 1992; Majumdar *et al.*, 1993; Weinstein and Majumdar, 1994). This suggests that the fractal dimension may likewise be a promising parameter for predicting mechanical properties. Addressing this question is therefore a major goal of the current research. For completeness and reference with other studies, additional microstructural measures to be included in correlation studies are the wet and dry densities and the ash weights.

As mentioned previously and indicated in the above outline, the research is being carried out at several different laboratories. The Orthopedic Biomechanics Laboratory at the University of Texas Medical School is under the direction of Dr. Timothy Harrigan and also staffed by Dr. Catherine Ambrose and Ms. Frances Biegler. This facility provides the source for fresh human bones and also is equipped for specimen cutting and mechanical testing. A recently funded collaborative research project between the UT Medical School, the Space Biomedical Research Institute at NASA Johnson Space Center, and the Baylor College of Medicine provides the focal point and programmatic framework for the current research. Drs. Linda Shackelford and Laurie Webster oversee the NASA participation providing guidance and direction on scientific issues as well as maintaining the relevance of the work to NASA programs and mission. Drs. Adrian LeBlanc, Harlan Evans, and Chen Lin of the Baylor College of Medicine are responsible for all MRI scanning and image preparation. Fractal analysis is being conducted by Dr. Raj Acharya of the Biomedical Imaging Group at the State University of New York at Buffalo. As a result of participation in the Summer Faculty Fellowship program the tasks associated with specimen preparation, mechanical testing, and post-test measurement of physical properties will likely be shared between the UT Medical School and facilities at Texas A&M University.

More detailed descriptions of the research activities are outlined below. The overall project can be roughly divided into four phases based upon the size and nature of the bone specimens being examined. The first phase involves acquisition, preparation, and imaging of whole bone human tibia specimens. The second phase deals with preparation and imaging of a 3-4 cm section excised from the proximal tibia just beneath the tibial plateau. The third phase involves further cutting the specimen into slabs roughly 1 cm thick and imaging each slab. The fourth and final phase involves cutting the slabs into 1 cm cubes, mechanically testing the cubes, and analyzing the physical properties of the tissue in each cube specimen.

Phase 1 – Whole Tibia Procedures

Whole tibia bones are acquired through the University of Texas Medical School and prepared for imaging in the Orthopedic Biomechanics Laboratory. Each tibia is cleaned of soft tissue and stored frozen. The bone is first X-rayed to ensure that it is free of disease or otherwise not suitable to be included in the study. For proper MRI scanning, each bone must be mounted in a container such that the bone is immersed in 'doped' water. A container for this purpose was fabricated at NASA/JSC from 1/4" thick plexiglass plate material. The box assembly with a tibia mounted is depicted in Figure 1. The plexiglass box is 4.5" wide, 4" tall, 16.5" long, and open at the top. A lid was made to cover the box to minimize splashing or leakage of the fluid during handling. An insert piece was also made to actually hold the bone. This permits easy placement and removal of the bone from the box. It was also intended to provide easy removal of the bone after it is embedded in plastic subsequent to MRI scanning. The insert was constructed from 1/8" plexiglass and made to fit just inside the walls of the box like a thin liner. Holes (1/4" diameter) were drilled in the long sides of the insert for holding rods inserted transversely through the bone. Two transverse holes were drilled diametrically through the bone near the proximal and distal bounds of the diaphysis region. The holes are 1/4" in diameter and lie in the medial/lateral plane. Plexiglass rods were inserted through these holes and then mounted in the holes in the insert liner.

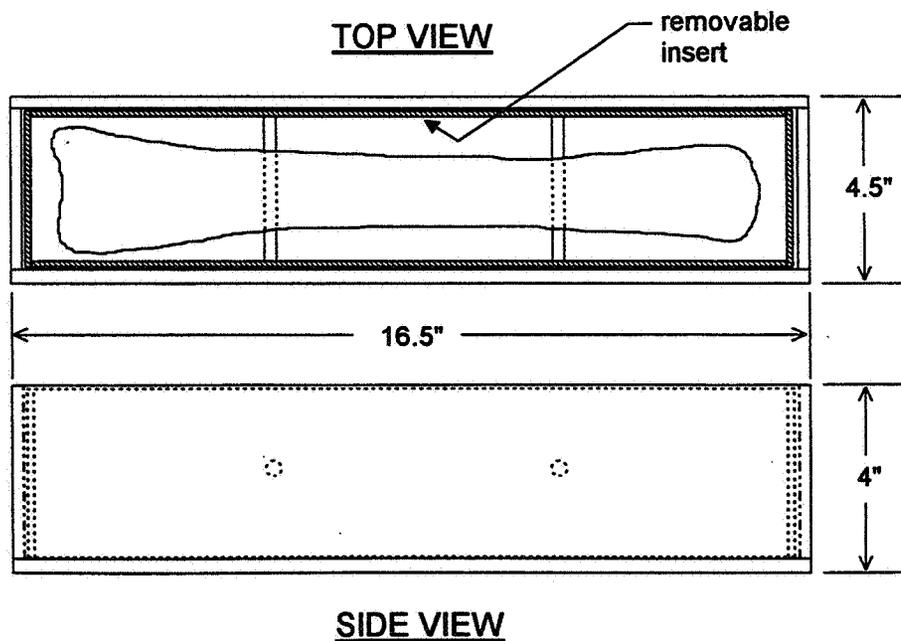


Figure 1.- Schematic drawing of container for whole bone MRI scanning.

The final preparation for imaging is to mark the boundaries of the proximal section from which mechanical testing specimens will be cut. This section should typically be roughly 3 to 4 cm long and located just beneath the tibial plateau. Landmarks must also be created to provide reference marks for locating tissue regions within the plane of each image slice. Four shallow "grooves" in the outer surface will be made for this purpose. These grooves will run axially along the length of the bone and will be cut with a Dremel tool to a depth of 1 to 2 mm. The grooves will be located along the medial, lateral, anterior, and posterior aspects of the periosteal surface of the bone cortex. The grooves will appear as notches in each planar transverse image. The intersection of lines drawn connecting the medial/lateral notches and the anterior/posterior notches will thereby form a systematically defined origin for coordinate locations within the plane of each image.

After the bone is properly marked and mounted in its container it is transferred to the Baylor College of Medicine for MRI scanning. The whole bone scans are made using a knee resonator coil, which has an inside diameter of 20.5 cm and an axial length of 26.5 cm. Axial scans (i.e. transverse to the bone axis) were made of the 3 to 4 cm region identified in the proximal portion with no gap between adjacent scans and with a scan thickness of 2 mm. These scans actually required multiple images to permit the determination of more detailed parameters such as T1, T2, and T2* in addition to the more routine spin-echo signals. The in-plane resolution of these images is approximately 0.5 mm. Additional spin-echo scans are made at 2 to 3 cm intervals moving proximally into the diaphysis region of the bone. These will be used for further studies on constructing finite element models from MRI data. In addition, five roughly evenly spaced coronal scans (i.e. parallel to the bone axis in the medial/lateral plane) were also made in order to register the axial position of the axial scans.

Following the MRI scanning, the container was emptied of water and returned with the bone still mounted to the UT Medical School. The container was then filled with polyester resin to completely embed the bone in plastic. A plastic sheet liner was placed in the container first to facilitate removal after the plastic hardened. The embedding procedure was intended to provide orthogonal reference surfaces for the whole assembly to guide subsequent cutting operations and also for defining common coordinate systems for locating positions within images and within the physical bone slab specimens. This particular procedure is currently being reviewed in lieu of using the "grooves" described above for the same purpose. Further preliminary studies are underway to address this.

Phase 2 – Proximal Tibia Section Procedures

In order to obtain images with high enough resolution to depict the microstructural architecture within trabecular bone tissue the size of the bone sample to be imaged must be reduced. The MRI facilities at Baylor College of Medicine are again used for this imaging. An orbit coil is used for high resolution MRI but its chamber is only 8 cm in

diameter. Thus, a smaller section from the proximal portion of the tibia was cut using a band saw at the UT Medical School. This piece was roughly 3.5 cm long with the first cut 1-2 cm below the tibial plateau (see Fig. 2). Before making these cuts, care must be taken to document the axial location of the cuts relative to a landmark that can be identified in the MRI. Two choices are apparent: the proximal-most tip of the articular surface, or the location of the proximal transverse rod upon which the bone is mounted. The first tibia that was used in preliminary tests was rather large and had to be trimmed significantly on the medial and lateral aspects as well in order to fit within the orbit coil. This also meant cutting away essentially all of the surrounding plastic in which the bone had been embedded. The large size of the bone also permitted cutting an additional 3.5 cm section below (or distal to) the first. This piece was small enough to fit within the orbit coil without further trimming. The in-plane resolution of the orbit coil is 0.125 microns for a scan thickness of 3 mm. Axial scans were made for the entire length of the piece with no gap between successive scans. Similarly, coronal scans were made of the same piece with no gap between scans. These scans were initially made on both pieces with the specimens simply wrapped with paraffin sheets. Upon reviewing the images, however, artifacts due to air pockets were observed, particularly in the central portions of the pieces where the porosity is high. Thus, the protocol has been revised to have the pieces fully hydrated before scanning and placed in a container of water during scanning. Placing the specimens in water will also ensure that the reference notches on the external surface of the cortical wall are clearly imaged.

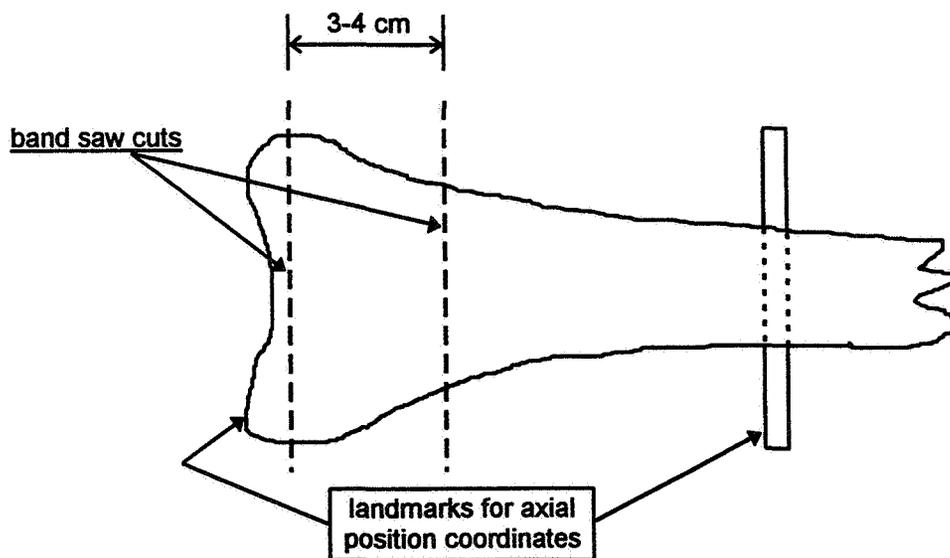


Figure 2.- Region cut from proximal tibia.

Phase 3 -- Transverse Slab Procedures

The next phase begins with cutting the 3 to 4 cm pieces into slabs roughly 1 cm thick or slightly thicker. The goal is to get 3 slabs from each piece. A band saw at the University of Texas Medical School was used for cutting the preliminary pieces examined thus far. Using a band saw makes it difficult to maintain precisely straight, flat, and parallel cutting surfaces and this must be considered in subsequent cutting operations for making cube specimens. Proper marking and/or measuring must be included at this stage in order to maintain the axial position of the cut faces relative to coordinate landmarks. The width of material removed by the saw must also be accounted for in this process. The next two steps involve imaging each separate slab and the order in which they are carried out is not critical. An X-ray image of each slab will be made by contact radiography at the UT Medical School. These X-rays will produce detailed images of the trabecular structure from which alternate calculations of fractal parameters can be made. This will provide a direct comparison of the image quality and fractal parameters between images derived from X-ray and those from MRI. The slabs will also be imaged using dual-energy X-ray absorptiometry (DEXA) scanning techniques at the Baylor College of Medicine. DEXA scans are available clinically and give measures of bone mineral density (BMD) and bone mineral content (BMC). The BMD and BMC values can be determined for regions of interest defined graphically on the image. Thus, with proper documentation of coordinates, BMD and BMC can be calculated for the same specific tissue regions from which the cube specimens for mechanical testing were ultimately cut.

Phase 4 -- Cube Specimen Procedures

The final phase of the protocol begins with cutting each slab into a series of roughly cube shaped specimens for mechanical testing. Considerable time and effort has been spent in developing and refining these procedures but only the highlights will be summarized here. First, each slab must be marked in some manner to identify which flat surfaces are proximal (or superior) and/or distal (or inferior) and which reference notches are medial/lateral and anterior/posterior. These markings should preserve these identifications to the degree possible throughout the process of being cut into the cube specimens. A Buehler Isomet low speed diamond blade wafering saw is used for this cutting. The saw used during the summer research period was located at the University of Texas Medical School, but one is also available at Texas A&M University. A series of parallel cuts are made to cut the slab first into "bars" as indicated by the long-dashed lines in Figure 3. Two blades were gang-mounted approximately 1 cm apart to create two parallel cuts with each cutting operation. A custom gripper was made to hold the slab during this process to allow 3 or 4 such bars to be cut without re-gripping the workpiece. The precision of these cuts will routinely produce surfaces parallel to well within required tolerances for opposite faces of the cube specimens. The surfaces of the bars formed from the band saw will not be parallel enough, however, so each bar must be rotated 90 degrees

and cut to face off one of the two surfaces cut with the band saw. Each bar is next cut transverse to its long axis to yield a set of cube specimens (these cuts are indicated by the short-dashed lines in Fig. 3). Each specimen must be properly marked to identify its orientation and coordinate location within the slab cross-section. A unique requirement of the current work is the need to cut slabs before making cross-cuts to produce cubes. This is necessary to allow DEXA and X-ray imaging as described previously and thereby precludes using a milling machine with blade cutters (as used by Ciarelli *et al.*, 1991, for example). The detailed steps and procedures in the cutting process have been developed with the assistance of a volunteer undergraduate research assistance and are outlined in a separate document resulting from that work (Brandt, 1995).

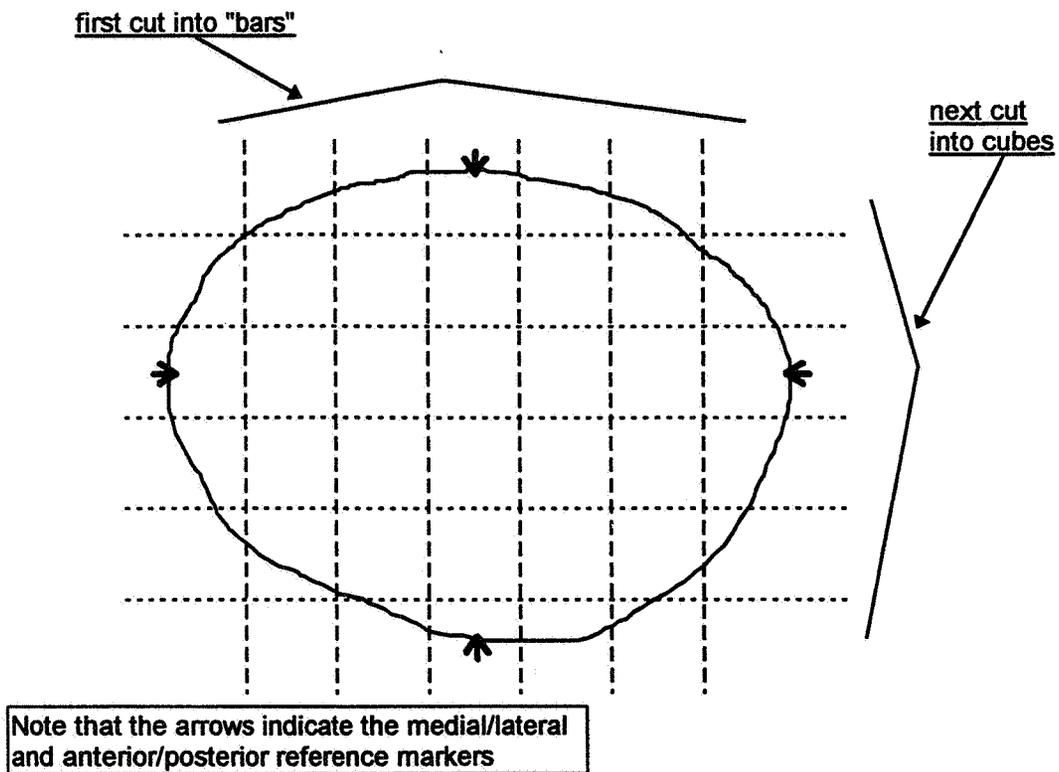


Figure 4.- Cutting slabs into cube specimens.

Cube specimens will be stored wet and frozen until time for mechanical testing. Mechanical tests will be conducted in compression under displacement control at a rate of 0.01s^{-1} . Each test specimen will be loaded between two flat platens lubricated to eliminate lateral constraint due to friction. One of the platens will be articulated to self-align and reduce non-uniform loading from imperfectly parallel specimen surfaces. A shallow (1-2 mm deep) recessed area will be machined in the platens to facilitate placement of the specimens centered along the machine loading axis. Each specimen will be tested nondestructively along each of its 3 orthogonal axes (i.e., superior/inferior, medial/lateral,

anterior/posterior) following procedures similar to those of Goulet et al. (1994), Keller (1994), and Linde et al. (1992). This will permit calculation of elastic moduli for each direction. A final destructive test to failure will then be conducted along the primary anatomical axis of loading (superior/inferior for proximal tibial bone). Quantities such as the ultimate strength, ultimate strain, and energy absorbed to failure can then be calculated for this direction of loading. Additional details to be determined through preliminary testing and further study include whether to use preconditioning cyclic loading, which particular testing machine to use, and whether to use extensometers for surface strain measurement. Using extensometers will likely require more elaborate design of the platens in order to provide adequate clearance to prevent the extensometers from interfering with the platens during testing. Numerous studies have addressed the issue of end effects (Allard & Ashman, 1991; Aspden, 1990; Harrigan *et al.*, 1988; Keaveny *et al.*, 1993; Linde *et al.*, 1992; Simmons & Hipp, 1995; Zhu *et al.*, 1994), but no simple solution has been developed to date, including using extensometers. Thus, applying an extensometer will definitely provide more insight into the details of each test, but it will not totally mitigate the problems associated with end effects. Testing machines are available at the University of Texas Medical School and also at Texas A&M University. Facilities are also available at Texas A&M for determining the wet and dry densities and ash weights. The wet density is simply the wet weight of the specimen (in g), which can be measured before or after testing, divided by the volume of the specimen (in cm^3) as calculated from caliper measurements of physical dimensions. The dry density is the weight of the specimen (in g) taken after drying in an oven at 100°C for 24 hours. The specimen is then ashed in an oven at 500°C for 48 hours and weighed. The ratio of ash weight (in g) to dry weight (in g) is commonly expressed as a percentage and termed the "ash weight percent", or sometimes even just "ash weight". An ash density can also be calculated by dividing the ash weight (in g) by the volume (in cm^3).

SUMMARY

Much progress has been made during the summer research period in defining the relevant requirements, evaluating available options, and establishing detailed procedures for conducting the tests and analyses required for this complex, multi-disciplinary, multi-institution collaborative research effort. Extensive study of the literature has been combined with preliminary tests as needed to address the major issues encountered. Four tibia (2 sets of paired) have been acquired and mounted in the removable insert pieces for whole bone MRI scanning. One has already been scanned and used for the preliminary studies. As the preliminary work progressed through mechanical testing the images from the MRI scans are being reviewed and analyzed by Dr. Acharya. When all details of the protocol are approved and agreed upon by all investigators, the other three tibia will be tested. Correlation studies will then be conducted to establish the relationships between the measured mechanical properties and the various independent variables (fractal dimension, wet/dry/ash densities, BMD/ BMC, $T2^*$, etc.).

REFERENCES

- Acharya, R. S., LeBlanc, A., Shackelford, L. C., Swarnarkar, V., Krishnamurthy, R., Hausman, E., and Lin, C. (1995) Fractal analysis of bone structure with applications to osteoporosis and microgravity effects. (manuscript submitted).
- Allard, R. N. and Ashman, R. B. (1991) A comparison between cancellous bone compressive moduli determined from surface strain and total specimen deflection. *Trans. Orthop. Res. Society 37th Annual Mtg.*, 151.
- Aspden, R. M. (1990) The effect of boundary conditions on the results of mechanical tests. *J. Biomechanics* **23**, 623.
- Brandt, S. (1995) Procedures for cutting cube specimens of trabecular bone from the proximal tibia. Unpublished Research Report.
- Ciarelli, M. J., Goldstein, S. A., Kuhn, J. L., Cody, D. D., and Brown, M. B. (1991) Evaluation of orthogonal mechanical properties and density of human trabecular bone from the major metaphyseal regions with materials testing and computed tomography. *J. Orthop. Res.* **9**, 674-682.
- Figueroa, F. (1995) Methodologies to determine forces on bones and muscles of body segments during exercise, employing compact sensors suitable for use in crowded space vehicles. *Final Report NASA/ASEE Summer Faculty Fellowship Program*.
- Goulet, R. W., Goldstein, S. A., Ciarelli, M. J., Kuhn, J. L., Brown, M. B., and Feldkamp, L. A. (1994) The relationship between the structural and orthogonal compressive properties of trabecular bone. *J. Biomechanics* **27**, 375-389.
- Harrigan, T. P., Jasty, M., Mann, R. W., and Harris, W. H. (1988) Limitations of the continuum assumption in cancellous bone. *J. Biomechanics* **21**, 269-275.
- Keaveny, T. M., Borchers, R. E., Gibson, L. J., and Hayes, W. C. (1993) Trabecular bone modulus and strength can depend on specimen geometry. *J. Biomechanics* **26**, 991-1000.
- Keller, T. S. (1994) Predicting the compressive mechanical behavior of bone. *J. Biomechanics* **27**, 1159-1168.
- Linde, F., Hvid, I., and Madsen, F. (1992) The effect of specimen geometry on the mechanical behaviour of trabecular bone specimens. *J. Biomechanics* **25**, 359-368.

- Majumdar, S., Weinstein, R. S., and Prasad, R. R. (1993) Application of fractal geometry techniques to the study of trabecular bone. *Med. Phys.* **20**, 1611-1619.
- Ruttiman, U. E., Webber, R. L., and Hazelrig, J. B. (1992) Fractal dimension from radiographs of peridental alveolar bone. *Oral Surg.Oral Med.Oral Pathol.* **74**, 98-110.
- Simmons, C. A. and Hipp, J. A. (1995) A new method for the mechanical testing of trabecular bone. *Proc. 1995 ASME Bioengineering Conf.*, **BED-Vol. 29**, 355-356.
- Todd, B. A. (1994) Finite element modeling of the lower extremities. *Final Report NASA/ASEE Summer Faculty Fellowship Program.*
- Weinstein, R. S. and Majumdar, S. (1994) Fractal geometry and vertebral compression fractures. *J. Bone Min. Res.* **9**, 1797-1802.
- Zhu, M., Keller, T. S., and Spengler, D. M. (1994) Effects of specimen load-bearing and free surface layers on the compressive mechanical properties of cellular materials. *J. Biomechanics* **27**, 57-66.

CONSTRAINTS IN GENETIC PROGRAMMING

Final Report

NASA/ASEE Summer Faculty Fellowship Program – 1995
Johnson Space Center

Prepared By: Cezary Z. Janikow, Ph.D.
Academic Rank: Assistant Professor
University & Department: University of Missouri – St. Louis
Department of Mathematics and
Computer Science
St. Louis, MO 63121

NASA/JSC
Directorate: Engineering
Division: Automation, Robotics and Simulation
Branch: Intelligent Systems
JSC Colleague: Dennis Lawler
Date Submitted: July 20, 1995
Contract Number: NGT-44-001-800

ABSTRACT

Genetic programming refers to a class of genetic algorithms utilizing generic representation in the form of program trees. For a particular application, one needs to provide the set of functions, whose compositions determine the space of program structures being evolved, and the set of terminals, which determine the space of specific instances of those programs. The algorithm searches the space for the best program for a given problem, applying evolutionary mechanisms borrowed from nature.

Genetic algorithms have shown great capabilities in approximately solving optimization problems which could not be approximated or solved with other methods. Genetic programming extends their capabilities to deal with a broader variety of problems. However, it also extends the size of the search space, which often becomes too large to be effectively searched even by evolutionary methods. Therefore, our objective is to utilize problem constraints, if such can be identified, to restrict this space. In this publication, we propose a generic constraint specification language, powerful enough for a broad class of problem constraints. This language has two elements – one reduces only the number of program instances, the other reduces both the space of program structures as well as their instances. With this language, we define the minimal set of complete constraints, and a set of operators guaranteeing offspring validity from valid parents. We also show that these operators are not less efficient than the standard genetic programming operators if one preprocesses the constraints – the necessary mechanisms are identified.

INTRODUCTION

Evolutionary problem solving simulates nature to search for a solution. First, representation for potential solutions must be defined, along with an evaluation function to quantify different solutions. A number of solutions are simultaneously processed in a simulated population. Individual solutions undergo simulated evolution: Darwinian selective pressure is used for survival determination, mutation is used to alter individual solutions, and crossover is used for information inheritance from usually two parents to one or two offspring.

Genetic programming (GP) [5, 6] is one of the most recent evolutionary methods. It differs from others, such as the well-known genetic algorithms (GAs) [2, 3], by its representation – an individual solution is a high level program, structured as a tree representing the dynamic program structure. This allows, for example, for the learning of analytical functional descriptions [5] – which can not be accomplished with GAs.

A GP application requires specification of the primitive function set and the terminals, which can be used in any combination (the *closure* property) [5]. However, by providing the primitive functions (along with procedural interpretations), one explicitly determines the set of plausible expressions that can be evolved. For example, suppose we need to learn the (unknown) function $y = \sin(x) \cdot \cos(x + 2)$. If we don't know the function, we do not know that the solution will involve trigonometric functions. Therefore, we may decide to use the following function set $\{+, -, *, /\}$. This will prevent the exact solution from being discovered (the *sufficiency* principle states that one assumes that the function set includes all needed functions). The evolved solution would have to be fairly complex for a satisfactory approximation (if possible at all, especially given size constraints imposed on GP programs). On the other hand, if we allow many functions to participate in the primitive set, we explode the search space beyond manageability. Therefore, the GP approach is very sensitive to the user's insights (in addition to being very sensitive to its own parameters). This will hopefully change when methods are developed to constrain the space, to incorporate heuristics, to automatically select/prune the set of primitives, to automatically update the sensitive parameters, *etc.*

An important issue for any problem solving method is that of handling constraints, which are often present. If disregarded, they may lead to infeasible solutions. On the other hand, when properly handled they can reduce the amount of searching required. Constraints can be handled by evolutionary algorithms, especially by genetic algorithms, where most constraint-related research has been done. However, a common problem is that of generality of any approach. Many GA approaches create specialized representations and/or operators, which prohibit invalid solutions from occurring. Examples are a matrix representation for the transportation problem and a permutation sequence for the traveling salesman problem. Even though these approaches are very nicely crafted and are efficient, they are also hand-tailored for the specific problems and must be redone for a new problem, a class of problems, or a specific instance of a problem with different constraints. To avoid that, one needs to provide a more general approach. One such approach has been studied in GAs – penalize solutions which violate the constraints [8]. But this method is not perfect either. Even though it is generic and only requires modification of the evaluation function, it is much less efficient as it still allows explorations of infeasible problem subspaces, wasting resources there. In addition, too relaxed a penalty can still allow generation of infeasible solutions.

Too rigid a penalty may prohibit good solutions from being discovered [7].

The most desirable approach is to provide a constraint-specification language, and then provide built-in mechanisms to handle those constraints, preferably by moving the search into feasible subspaces only. The language should be powerful enough to express a broad range of specific constraints that a particular problem may have. This has been done with GAs for the class of linear constraints over a parameter space [7].

GP offers much greater capabilities than GAs by its variable length parameterized representation. However, as mentioned, it must advance in many directions to enjoy more practical applications. And constraints are an important aspect of such advancement. In this paper, we propose a constraint handling methodology, which is based on the idea of a constraint specification language. This language is presented and enforcement mechanisms are provided. The language is not capable of expressing "any" constraint. However, it is applicable to a broad range of problem constraints. Moreover, we show that the enforcement mechanisms do not reduce the efficiency of the GP algorithm. In fact, the actual search efficiency is greatly improved since the search is now conducted only in the feasible subspace.

The idea is based on restricting both the space of program structures, and on reducing the number of program instances for any particular program structure. In this paper, we overview the various methods for processing constraints in genetic algorithms. Then, we propose a constraint specification language for GP, which is easy to use. Afterwards, we present transformations aimed at reducing the set of constraint specifications to a minimal yet sufficient set, which is easy to enforce. Subsequently, we define mutation and crossover to be "closed" in the program subspace specified by the specifications. That is, given valid parents these operators generate only valid offspring. We also provide a method to initialize the population of programs with only valid instances to ensure that the evolution will be closed in the feasible subspace. Finally, we show that one may implement these operators with the same level of efficiency as the standard operators in unconstrained GP.

CONSTRAINTS IN GP AND GAs

Genetic programming is a special case (or a generalization, depending on the exact definition) of genetic algorithms. Given a problem, a set of constants for the problem (*e.g.*, program variables or actions) is specified, and a set of primitive functions is defined. Eventual program data can be included in the terminal set, or it can be hidden in interpretations of functions. Compositions of functions determine the space of programs which can be expressed this way. What is sought is one of such programs, the one which solves the problem the best way (given some criteria).

Discovery, or rather creation, of the best program is accomplished by evolution. A population of initial programs (possibly random) is set, which then evolve by simulating nature. In this simulation, current programs are evaluated (given problem criteria) and an evaluation-based pressure is used to promote survival of better programs. Programs also undergo mutation, and create offspring by means of crossover.

An obvious problem that GP tries to address is that of searching the infinite space of possible programs. This is done by limiting this space as explained below. Our objective is to further limit this space by additional constraints.

Current State of the Art in Constraint Processing in GP

Except for ADFs (automatic function definitions) and a few tailored applications, there is very little reported work on constraints and methods for handling them in GP [4, 5, 6].

In ADFs, constraints relate to differentiating among different functions and the program being evolved. This has few implications for our objectives. The tailored applications are not more helpful, since they aim at satisfying constraints of very specific functions (usually a single constraint).

In general, the only other constraint used in GP relate to managing the complexity of the evolution. This is done by imposing restrictions on program size. Program size is defined in one of two ways: either by the depth of the tree (proportional to maximum number of nested function calls) or by the total size of the tree (proportional to the total number of function calls). Either case is always handled in one of two different ways:

- abandon the violating program, and keep creating new ones until one that satisfies the constraint is generated (method 1 below),
- use the parent program instead of the violating offspring (method 2 below).

These methods are fine if violations are rare. With the size constraint, this is indeed the case. However, with new constraints, this is unlikely to be true and new methods will have to be investigated.

Various Methods for Handling Constraints

In any evolutionary method (such as GP, but especially in GA), constraints have been, or at least can be, handled in a number of ways. Difficulties come in implementing some of the ways, and in prior determination of which way might be the best for a given problem (or a class of problems). Unfortunately, there is no "silver bullet".

Genetic algorithms enjoy richer methodologies to deal with problem constraints. In addition to those two methods listed for dealing with size constraints in GP, GA researchers have used approaches such as modifications of the interpretation of a particular solution (called *phenotype*), repair algorithms, and penalty functions to penalize invalid solutions. A recently emerging approach is to provide "smart" operators, called *closed*, which are intended to produce only valid offspring from valid parents. Such operators have been proposed for GA parameter optimization with linear constraints [7].

Objectives and Methods

Because possible constraints are endless, it is impossible to provide constraint handling methods for specific constraints ahead of time. On the other hand, we want to avoid having to provide such methods separately for each problem, whenever a constraint appears. Therefore, our objective is to design a constraint specification language able to express constraints encountered in a broad range of applications. Then, we could have generic mechanisms processing those constraints so that

- only valid programs are evolved,
- the space of valid programs is constrained to reduce the effective search space.

EXPLORATION OF CONSTRAINT SPECIFICATIONS

Search Space in GP

GP's problem space is defined by terminals T and functions F , where each $f_i \in F$ has a fixed arity a_i . Function compositions determine the space of possible program structures. Because the space is infinite, it is normally restricted by the aforementioned two methods – restrict the number of nodes or restrict the maximal depth of a program. Terminals are values which can appear in terminal nodes. They have no implications for program structures. Instead, given a specific program structure, which serves as a template, terminals determine the space of possible program instances. This space, for a particular program structure, can again be infinite if infinite sets are allowed (such as real numbers). Therefore, the space of all programs is determined by both F and T , and this space can be infinite in two dimensions. One of these possibly infinite dimensions, the space of program structures, is normally restricted by the size constraints. The other dimension, program instances of a program structure, is restricted by the finite accuracy of computer representations. It is assumed that this space contains the sought program. This assumption is the basis for GP problem solving – the *sufficiency* assumption.

T - and F - Constraint Specifications

We use two different kinds of constraint specifications, but, as we will see, they are not completely independent. These are syntactic T -specifications and semantic F -specifications. Moreover, we will show that not all of them are really needed – after certain transformations, only a few are sufficient to express exactly the same constraints. However, it will generally be much easier to express problem constraints with the full specification language, while it will be much easier to devise mechanisms for the minimal set.

T -specifications are based on domains for function arguments and on function ranges.

Definition 1 Let us define \Rightarrow to stand for domain compatibility. That is, $X \Rightarrow Y$ means that X can replace Y , where both X and Y stand for sets of values (finite or infinite) allowed for domains or returned as function ranges.

Definition 2 Define the following T -specifications (syntactic constraints):

1. T^{Root} – the set of values allowed at the Root, or the set of values allowed to be returned by the evolved program, that is by functions appearing at the Root. T^{Root} actually specifies both a domain (for the root node) and a range (for the program).
2. $T_* - T_i$ – T_i is the range of f_i , that is the set of values returned by the function f_i .
3. $T_*^j - T_i^j$ – T_i^j is the domain for the j^{th} argument of f_i , that is the set of values allowed there (which may be returned by functions used as this argument).
4. $T_* \overset{?}{\Rightarrow} T_*^*$ – compatibilities between ranges and domains
5. $T_* \overset{?}{\Rightarrow} T^{Root}$ – compatibilities between function ranges and the program range

where '?' indicates that the compatibility can be straight or it can be negated.

T-specifications reduce both the space of program structures and the space of instances of those structures. Therefore, they are very powerful constraints. GP uses only one terminal set – any value from that set may appear in any leaf of a program tree. Obviously, this is not true in actual programs – different functions take different arguments. *T-specifications* allow expressing such differences, thus allowing reduction in the space of program instances per program structure. Moreover, some *T-specifications* also implicitly restrict what function can call other functions, effectively reducing the space of possible program structures. For example, if the domain for a function argument is fixed, then the value for that argument may not be obtained from a function with an incompatible range. Therefore, some *T-specifications* can be seen analogous to *function prototypes* available in high level languages.

However, syntactic fit does not necessarily mean that a function *should* call another function. One needs additional specifications based on program *semantics*. These are provided by means of *F-specifications*, which further restrict the space of program structures (but not program instances).

Definition 3 Define the following *F-specifications* (semantic constraints):

6. F^{Root} – the set of functions disallowed at the Root.
7. F_* — F_i is the set of functions disallowed as direct callers to f_i (generally, a function is unaware of the caller; however, GP constructs a program tree, which represents the dynamic structure of the program).
8. F_i^* — F_i^j is the set of functions disallowed as arg_j to f_i .

Example 1 Assume a function (if arg_1 arg_2 arg_3), interpreted as: if arg_1 evaluates to true, return the evaluation of arg_2 , else return the evaluation of arg_3 (assume both of which evaluate to real numbers). One needs to specify that arg_1 could only be terminals which are boolean values, or only functions which return boolean values. Assuming that $T = R^1 \cup \{T, F\}$, one may specify $T_{if}^1 = \{T, F\}$. Because R^1 is not compatible with $\{T, F\}$, only elements of the latter can be placed there.

Proposition 1 $X \Rightarrow Y \leftrightarrow X \subseteq Y$.

$\therefore X \Rightarrow Y$ means that in places where values from Y are valid one may place any value from X , or any function returning a value from X . To guarantee that no out-of-domain values are used for the original Y , X may not contain values not found in Y . Therefore, it must be a subset of Y , or it must equal Y .

Using known properties of \subseteq , domain compatibilities could be automatically computed (giving compatibility *T-specifications* #4 & #5), as long as these are restricted to syntactic constraints.

Example 2 Assume two sets: $T_1 = \{1, 2, 3\}$ representing masses of physical objects in kilograms, and $T_2 = \{1, 2\}$ representing times in seconds. Thus one may conclude that $T_2 \Rightarrow T_1$ since $\{1, 2\} \subset \{1, 2, 3\}$. But by observing the interpretations of these objects, an obvious conclusion is that $T_2 \not\Rightarrow T_1$, but this is based on interpretation of these sets, which is left to *F-specifications*.

Rules on T - and F -specifications

Given the above T -specifications and F -specifications, which can be used to express problem constraints, an obvious issue is that of possible redundancies, or that of existence of sufficiently minimal specifications. We answer these questions in this section. Surprisingly, after certain static transformations, only a subset of T - and F -specifications will turn out to be sufficient to express all T - and F -specification constraints. This observation is extremely important, as it will allow efficient constraint enforcement mechanisms after initial preprocessing. The sufficient minimal set is thus important for efficient constraint processing, but not for constraint specifications – specifications are more easily expressed with the original T -specifications and F -specifications. This is why we need both, along with the necessary transformations.

The first step is to extend F -specifications.

Definition 4 Define 'complete' T -specifications as those that list all elements of Definition 2, including ranges and domains for all functions and their arguments and compatibilities between all pairs range–domain and range–program range.

Proposition 2 The following F -specification constraints are implied by complete T -specifications:

$$\begin{aligned} \forall_{f_k \in F} (T_k \not\# T_i^j \rightarrow f_k \in F_i^j) \\ \forall_{f_k \in F} (T_k \not\# T^{Root} \rightarrow f_k \in F^{Root}) \end{aligned}$$

:: If f_k returns a range which is not compatible with the domain for a specific function argument, then f_k cannot be used to provide values for the argument. The same applies to values returned from the program.

Proposition 2 is very important because the compatibility T -specifications (#4 & #5) can be automatically generated from other T -specifications, and according to the rule, they can be automatically translated to F -specifications. The latter, as we will see, are easier to handle.

Note that the opposite of these implications is not true since some F -specifications are based solely on interpretations. In other words, it is not true that $f_k \in F_i^j \rightarrow T_k \not\# T_i^j$. Note that the following is not true either: $\forall_{f_k \in F} (T_k \Rightarrow T_i^j \rightarrow f_k \notin F_i^j)$ (see Example 2). Fortunately, the first implication is sufficient for us as it tells us that properly extended F^* and F^{Root} specifications subsume the $T_* \not\# T_*^*$ and $T_* \not\# T^{Root}$ T -specifications.

Example 3 Suppose $f = \{f_1, f_2\}$, $a_1 = 1, a_2 = 1$. Also suppose f_1 returns real-valued numbers ($T_1 = R^1$), and f_2 takes boolean arguments ($T_2^1 = \{F, T\}$). Because $T_1 \not\# T_2$, we may conclude that f_1 cannot be placed as the argument to f_2 : $f_1 \in F_2^1$.

Definition 5 If F -specifications explicitly satisfy Proposition 2 then call them 'T-extensive' F -specifications. If F -specifications do not explicitly satisfy Proposition 2 for any function $f_k \in F$, then call them 'T-intensive' F -specifications.

In other words, T -intensive F -specifications list only some additional constraints – which cannot be derived from T -specifications. T -extensive F -specifications, on the other hand, are those semantics-based constraints extended by syntactic constraints on function calls.

For now, we will look at redundancies among F -specifications.

Proposition 3 Suppose $f_k \in F$ and F -specifications are T -extensive. Then

$$\forall f_i \in F (f_k \in F_i \leftrightarrow \forall j \in [1, a_k] f_i \in F_k^j)$$

:: If a function f_i cannot call f_k , then f_k will never be called by f_i . Also, if f_k is never called from f_i , it must not be called from any of f_i 's arguments.

With Proposition 3 one may wonder whether we need both F_*^* and F_* constraints – they seem equivalent. The next rule says they are not.

Proposition 4 Suppose $f_k \in F$ and F -specifications are T -extensive. Then

$$\forall f_i \in F (\exists j \in [1, a_i] f_k \in F_i^j \not\rightarrow f_i \in F_k)$$

The implication is true only when Proposition 3 applies.

:: If f_k cannot be called from f_i by its j^{th} argument, it may possibly be allowed as another argument (unless, according to Proposition 3, it cannot be called from any of the arguments).

Even though they are not equivalent, both are not needed either. It turns out that F_*^* F -specifications are stronger.

Definition 6 If F -specifications explicitly satisfy Proposition 3, call them ' F -extensive' F -specifications. If F -specifications do not include any F_* constraints, call them ' F -intensive' F -specifications.

Proposition 5 F -intensive F -specifications are sufficient to express all possible F -specifications.

:: According to Proposition 3, $f_k \in F_i$ can be deduced when f_i is excluded from all arguments of f_k . According to Proposition 4, it can happen only when Proposition 3 applies. Therefore, F -intensive F -specifications provide sufficient information to produce F -extensive F -specifications.

We now return to the question of T -specifications vs. F -specifications. We have seen that T -intensive F -specifications provide restrictions on function calls based on interpretations, and that they can be extended to T -extensive F -specifications, which also take syntax into account. One question that comes to mind is: do we still need T -specifications after they have been used to produce T -extensive F -specifications? In other words, is there any constraint in T -specifications which is not expressed with T -extensive F -specifications? The answer is 'no' for certain T -specifications.

Proposition 6 T -extensive F -specifications are sufficient to express constraints imposed by compatibility (#4 & #5) and T_* (#2) T -specifications.

:: Let us look at compatibilities of the form $T_k \xrightarrow{3} T_i^j$. Proposition 2 says that the negated forms ($\not\Rightarrow$) are all expressed in T -extensive F -specifications. However, the straight form (\Rightarrow) can be superseded by F -specifications, which provide additional constraints based on interpretations. Thus, if $f_k \in F_i^j$, then the corresponding T -specification is irrelevant. On the other hand, if $f_k \notin F_i^j$ (in the T -intensive form), then we have two cases:

- if $T_k \not\Rightarrow T_i^j$, then according to Proposition 2 we put f_k into F_i^j : $f_k \in F_i^j$ in T -extensive forms
- if $T_k \Rightarrow T_i^j$, then we have no reason to extend F -specifications— thus, $f_k \notin F_i^j$

The same can be argued for $T_k \stackrel{?}{\Rightarrow} T^{Root}$. As to T_* T -specifications, they are sets of values returned by functions. Therefore, they place restrictions on function calls. But, F -extensive F -specifications express all possible restrictions on function calls. Said differently, T_* is only used for other specifications.

Definition 7 Define T -extensive F -intensive F -specifications as the 'normal' form.

Theorem 1 (*Fundamentals of T - and F -specification constraints*) Even if the user provides only T - F -intensive F -specifications, T - F -extensive F -specifications can be computed, and along with domains and the program range they are sufficient to express all T - and F -specification constraints. Moreover, just the normal F -specifications along with domains and the program range are sufficient as well.
 \therefore It follows from Propositions 5 and 6.

Based on Theorem 1, we may now restrict our discussion to F -specifications only, assuming that these are in the *normal* form. To make sure they are, a simple preprocessing mechanism suffices.

EXPLORATION OF CONSTRAINT HANDLING METHODS

We propose to implement the specified constraints into "smart" operators. To do so, we must define operators "closed" in the valid program structure – from valid parents always generate valid offspring. This will also require an initialization procedure with valid programs.

Definition 8 In the program tree, we call 'function nodes' all nodes which correspond to a function. In this case, we say that the function labels the node. All other nodes are called 'terminal nodes'.

Definition 9 Define \mathcal{T}_N to be the set of values which can replace node N . That is, \mathcal{T}_N is the set of values that the node can assume without invalidating, w/respect to T -specifications and F -specifications, the program tree containing that node.

Definition 10 Define \mathcal{F}_N to be the set of functions which can replace node N . That is, \mathcal{F}_N is the set of functions which can label that node without invalidating, w/respect to T -specifications and F -specifications, the program tree containing that node.

For terminal nodes, we cannot determine what other possible values can it contain by just looking at the node. We must look at the parent of the node (unless it is the *Root*). For function nodes, we could either use the set of values returned by the function labeling that node (T_i for f_i). However, after replacing the function node with a terminal node, we

would anyway have to look at the context where the node appears. Therefore, we decided to use the context information even for function nodes.

As the subsequent rules state, the above sets not only can be efficiently computed, but some can also be guaranteed to be non-empty under certain conditions, which hold for GP. Moreover, in the next section we will see that these sets can be precomputed for all possible node types, and that functions to extract random elements of these sets can be precomputed as well. This will lead to a very efficient enforcement of these constraints.

Proposition 7 *Assume a node N is the j^{th} argument of f_i and F -specifications are normal. Then,*

$$\mathcal{T}_N = T_i^j$$

$$\mathcal{F}_N = \{f_k | (f_k \in F) \wedge (f_k \notin F_i^j)\}$$

:: Any value that does not invalidate the domain T_i^j is OK. Any function that is not explicitly excluded from F_i^j is OK. This is so because if $f_i \in F_k$, that is if f_i cannot be accepted as a caller to f_k , then according to Proposition 3 $f_k \in F_i^j$, but it is not.

Proposition 8 *Assume a node N is the Root and F -specifications are normal. Then,*

$$\mathcal{T}_N = T^{\text{Root}}$$

$$\mathcal{F}_N = \{f_k | (f_k \in F) \wedge (f_k \notin F^{\text{Root}})\}$$

:: Arguments are analogous to those for Proposition 7, except that the Root provides the constraints.

Proposition 9 *$\mathcal{T}_N \neq \emptyset$ for any terminal node in any valid program.*

:: The valid program does not change when the terminal node is replaced with itself.

Proposition 10 *As long as any function returns a value (as it is in GP), $\mathcal{T}_N \neq \emptyset$ for any function node in any valid program.*

:: If the function node is labeled with f_i , then it can be replaced with any terminal from T_i . This set is not empty as long as each function returns at least one value.

Proposition 11 *For any function node N of any valid program, $\mathcal{F}_N \neq \emptyset$.*

:: If the node is labeled with f_i , then $f_i \in \mathcal{F}_N$.

Note that \mathcal{F}_N is not guaranteed to be non-empty for terminal nodes. That is, some terminals may only be used for computations, but will never be computed.

Example 4 *Suppose $F = \{f_{\square}\}$, and f_{\square} returns the closest integer to its real-valued argument. Then, $T_{\square} = I$, $T_{\square}^1 = R^1$, and $T^{\text{Root}} = I$. Also, $(f_{\square} \in F_{\square}^1) \wedge (f_{\square} \in F_{\square})$ (in the T - F -extensive form), but only $(f_{\square} \in F_{\square}^1)$ is sufficient (in the normal form).*

For the program $(f_{\square} \text{ 3.27})$, the terminal node 3.27 has $\mathcal{T} = I$ and $\mathcal{F} = \emptyset$.

We can now define closed operators. We assume that all random numbers are taken from a uniform distribution. For any node N , denote

- r_N^T to be a random element from \mathcal{T}_N
- r_N^F to be a random element from \mathcal{F}_N (assuming that it is non-empty)

For any terminal node N , denote

- v_N to be the current value from that node

Mutation

Assume that node N is chosen for mutation. This selection can be based on a fixed probability of mutating any *allele* in all chromosomes (often called *post mutation* in GP), or on selecting a random *allele* in a given selected parent (*normal mutation* in GP).

Operator 1 (mutation) *If a node N is selected for mutation, then replace it with r_N^T with probability p_m^1 , or with r_N^F with $1 - p_m^1$. If r_N^F is used, then recursively repeat exactly the same Operator 1 on all arguments of the selected function.*

If r_N^F is needed for the node N and the \mathcal{F}_N set is empty, try another random node from the same parent (in normal mutation) or abandon the operation (in post mutation). If r_N^F is needed for any descendent of N and the \mathcal{F}_N set is empty, use r_N^T instead.

Proposition 12 *For any valid parent program, mutation Operator 1 is guaranteed to take place as long as $p_m^1 > 0$. For a function node, Operator 1 is guaranteed to take place immediately. Moreover, all T - and F -specification constraints are guaranteed to be preserved. \therefore The parent is valid. According to Propositions 9 and 10, the set \mathcal{T}^N is never empty. Therefore, as long as this set is allowed in mutation ($p_m^1 > 0$), mutation will eventually take place on any node. However, if N is a function node, then according to Propositions 10 and 11 both \mathcal{T}_N and \mathcal{F}_N are non-empty, so mutation will immediately take place regardless of p_m^1 . The mutation sets are computed based on normal F -specifications, which are sufficient according to Theorem 1.*

Crossover

Because crossover with two offspring can be accomplished with two crossover operations, each with one offspring, we will define crossover with one offspring. In unconstrained GP, there are no specific constraints. Therefore, crossover is reduced to finding two random crossover points. In constrained "smart" crossover, the choices of plausible crossover points can be highly reduced. Requiring that two offspring can be generated from the same two crossover points further reduces chances of finding such points, but can be done if necessary.

Definition 11 *Define $\mathcal{S}_{N,x}$ to be the set of nodes from parent _{x} which can replace a given node N selected for crossover.*

Proposition 13 *For crossover at node N_1 in parent₁, and another parent₂,*

$$\mathcal{S}_{N_1,2} = \left\{ N_2 \left| \begin{array}{ll} v_{N_2} \Rightarrow \mathcal{T}_{N_1} & \text{if } N_2 \text{ is a terminal node in parent}_2 \\ f_i \in \mathcal{F}_{N_1} & \text{if } N_2 \text{ is a function node labeled } f_i \text{ in parent}_2 \end{array} \right. \right\}$$

:: Crossover may be seen analogous to mutation – from parent₂ select those crossover nodes which are also allowed in mutation for N₁. And \mathcal{T}_N and \mathcal{F}_N are terminals and functions that the node N₁ can mutate into. If N₁ = Root then \mathcal{T}_{N_1} and \mathcal{F}_{N_1} are terminals and functions that the node Root₁ can mutate into.

Operator 2 (crossover) *If parent₁ and parent₂ are two crossover parents, select a random crossover point N₁ in parent₁, except that internal nodes have collective probability of p_c^1 and leaves have collective probability $1 - p_c^1$ (following standard GP practice of directing crossover into internal nodes). Based on whether N₁ is the Root, apply Proposition 13 to compute $\mathcal{S}_{N_1,2}$. If the set is not empty, then select a random node N₂ (leaves and internal nodes may be given distinct probabilities with p_c^1), and replace the subtree starting with N₁ with that starting with N₂. If the set is empty, try another crossover point N₁ from parent₁.*

Proposition 14 *For any two valid parents, Operator 2 is guaranteed to find valid crossover, and the operation will satisfy T- and F-specifications.*

:: Both parents are valid. Therefore, replacing them wholly will produce a valid offspring. Moreover, the offspring is created by replacing a subtree with another subtree from a set computed according to T- and F-specifications. Therefore, any offspring will satisfy these constraints.

Feasible Initialization Procedure

Operator 3 (initialization) *Initialize the population by growing chromosomes starting each with a random terminal node N such that $v_N \in \mathcal{T}_{Root}$, and then applying Operator 1 to that node.*

Proposition 15 *The above initialization routine Operator 3 will only generate individuals which are valid with respect T- and F-specification constraints.*

:: Operator 1 guarantees a valid offspring from a valid parent (Propositions 7, 8). And the initial terminal node is valid as the Root.

IMPLEMENTATION

Constraint preprocessing

We do not need terminals T to be explicitly given as in GP – T_* , T_*^* , and T^{Root} will determine individual sets. The preprocessing needed to ensure that *F-specifications* are *normal* and that our operators can apply can be described as follows:

1. Read T^{Root} , T_* ranges for functions of F and T_*^* domains for their arguments
2. Read *T-specification* compatibility constraints of the form $T_* \stackrel{?}{\Rightarrow} T_*^*$ and $T_* \stackrel{?}{\Rightarrow} T^{Root}$ (not necessary if computed automatically)
3. Read (at least *T-intensive F-intensive*) *F-specifications*

4. Compute normal F -specifications
5. Produce functions for r^T and r^F for all necessary sets.

Given this preprocessing mechanism, the defined operators can be used in any GP.

Implementation

Proposition 16 r^T and r^F can be precomputed, as part of the preprocessing mechanism, into functions returning random elements of those sets.

:: For mutation, we directly need r_N^T and r_N^F . For any mutation, it is determined in one step which of the two is needed. Based on whether N is the Root or not, Proposition 8 or 7 gives us exactly the sets from which the random element is selected. There is a fixed number of these sets: there are exactly $1 + \prod_{f_i \in F} (a_i)$ of each T and F sets. All F sets are always finite with up to $|F|$ elements, and T is either finite, or infinite when domains such as reals are used. Moreover, these sets never change as GP operates. For the F sets, the elements can be enumerated and r^F can be compiled into a function returning a random function from each enumerated set. For the T sets which are infinite, r^T can be precompiled to returning a random entry from the domain. For finite sets, the elements can be enumerated again and r^T can be compiled into a function returning a random element from each enumerated set. For sets which are unions of finite or infinite subsets, one may first determine which class of subsets to choose from (assuming that we provide some measures comparing cardinalities of finite and infinite sets, or the user provides such information), and then apply one of the two above techniques.

For crossover, we need to use the $S_{N,a}$ sets. However, at each time we know whether the node N_a is a terminal or a function node, at which moment the problem reduces to the same as in mutation – selecting random entries from the appropriate T or F set. Moreover, if P_c^1 is used, the elements may be divided into two groups from which to select the random entry – p_c^1 would determine which group to use.

Theorem 2 (Implementation theorem for GP) The defined mutation and crossover operations not observing size constraints are as efficient as the standard operators in GP, when implemented with the preprocessing mechanism.

:: In GP, mutation generates a random function from F or a random element of T . Crossover selects a random subtree. It follows directly from Proposition 16 that in our approach any mutation or crossover can be accomplished by selecting a random entry from a fixed set, even though the sets are more plentiful. However, for any node it is deterministic, in a fixed time, which set should be used.

Conjecture 1 Provided T -specifications and F -specifications are the maximal constraints that can be implemented into a generic constrain processing methodology in GP without invalidating Theorem 2.

:: Other constraints will require information about a node position in a tree – processing complexity would be a function of tree depth.

References

- [1] Lawrence Davis (ed.). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.
- [2] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, 1989.
- [3] Holland, J. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.
- [4] Kenneth E. Kinnear, Jr. (ed.) *Advances in Genetic Programming*. The MIT Press, 1994.
- [5] John R. Koza. *Genetic Programming*. The MIT Press, 1992.
- [6] John R. Koza. *Genetic Programming II*. The MIT Press, 1994.
- [7] Zbigniew Michalewicz & Cezary Z. Janikow. "GENOCOP: A Genetic Algorithm for Numerical Optimization Problems with Linear Constraints". To appear in *Communications of the ACM*.
- [8] J.T. Richardson, M.R. Palmer, G. Liepins & M. Hilliard. "Some Guidelines for Genetic Algorithms with Penalty Functions". In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann, 1989.

**ORBITER FLIGHT DECK REDESIGN
(A PHYSICAL LAYOUT FOR A FUTURISTIC FLIGHT DECK)**

**Final Report
NASA/ASEE Summer Faculty Fellowship Program--1995
Johnson Space Center**

Prepared By:	Mehrzad Khorsandi
Academic Rank:	Assistant Lecturer
University & Department:	Texas A&M University Civil Engineering Dept. Engineering Design Graphics College Station, TX. 77843
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Flight Crew Support
Branch:	Graphic Research and Analysis Facility (GRAF)
JSC Colleague:	James Maida
Date Submitted:	October 30, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

The purpose of this summer project was to develop a set of schematic drawings for redesign of the Space Shuttle flight deck from which a three dimensional computer drawings can be built and viewed in a virtual environment. In order to achieve this goal, first recommendations for overall redesign of Space Shuttle previously made by experts in the field were reviewed and relevant information were extracted and delineated. Original drawings of the Space Shuttle made by Rockwell were obtained and carefully examined. In order to implement and assess any modifications in terms of space saving parameters, it was determined that the drawings alone could not achieve this objective. As a complement, physical measurements of the mockup of Space Shuttle flight deck were made and the information was categorized and properly labeled on the original drawings. Then, space-saving redesign ideas, as motivated by expert recommendations on such things as information display panel upgrade by technologically advanced flat display units, were implemented. Next, the redesign ideas were executed on the Forward flight deck, Overhead Console, Right and Left Console, and Center Console. A new 3-D computer drawing of this was developed by modifying the existing drawing on the in-house developed software (PLAID). Finally, the drawing was transported to a Virtual Environment and observed.

INTRODUCTION

Background:

In order to keep Space Shuttle operating smoothly in the future, the current cabin, its equipments, and its operation should be redesigned. The new futuristic cabin which includes the latest and the most advanced technological equipments should help operate the Space Shuttle in a more suitable, reliable, economical, and safe manner. The Orbiter Advanced Cabin Design has been under consideration for many years and many studies have been done to satisfy this need. Some of the major issues in this relation are:

- 1) The cabin equipments and flight crew's use of them are archaic;
- 2) The cost of operation and maintenance through the use of the current obsolete equipment are considerably high;
- 3) Use of new, more advanced, and lighter weight equipment could help achieve cabin weight reduction allowing for higher payload to be hauled to the orbit;
- 4) Use of more advanced information systems could result in crew size reduction; and
- 5) The expected one billion dollar budget cut in the Space Shuttle program for 1996 fiscal year makes cost cutting measures more imperative.

The archaic equipment currently used in the Space Shuttle are heavy and occupied large volume as compared to the most advanced technological equipments. For example, the bulky Cathode Ray Tubes (CRT) used as display in the Forward flight deck can be replaced with modern light weight and flat display units. The focus of this Summer project was to redesign the interior of the flight deck to increase space and decrease weight, taking advantage of these technological advancements.

Objectives:

The first objective of this project was the preliminary design of the flight deck based on the recommendations previously made by the experts in the aeronautic and avionic field. The second objective was to prepare a computer model of the proposed changes to the flight deck which could be viewed in virtual environment for review and possible modifications.

EXPERTS RECOMMENDATIONS

To become familiar with the background of the flight deck redesign and arrive at the objectives of this project many NASA memos were provided which were mostly cost saving recommendations of the experts of the field for the space shuttle. Also many more articles and books were recommended as references. Even though some of them were very hard to find in short period of time but they were all very useful to the design process.

There were two types of expert recommendations provided along with the project description, to reduce the operating and maintenance cost of the Space Shuttle. The first type were general recommendations for a futuristic Space Shuttle as a whole. The second type were specific recommendations for a futuristic flight deck. A summary of these follows.

General Recommendations

These recommendations were made based on the goal of 50% reduction in the Space Shuttle maintenance and operating cost.

- 1) Use of integrated navigation system
- 2) Use of electronic integrated orbiter.
- 3) Use of Multifunction Electronics Display Subsystem.
- 4) Use of Fiber-Optics.
- 5) Use of Star Tracker Cameras
- 6) Use of Solid State Data Recorders.
- 7) Use of On-Board Automation.
- 8) Use of Electronic Voice Communication.
- 9) Crew size reduction.

Expert Recommendations Specifically for Flight Deck

- 1) Removal of the Center Console.
- 2) Removal of the two Side Consoles.
- 3) Removal of the Over Head Console.
- 4) Reduction of window acreage.
- 5) Flattening of the Forward Station (based on the use of the most advanced and less space taking equipment).
- 6) Flattening the AFT Station.
- 7) Cockpit Redesign / Flightcrew Escape.

THE DESIGN PROCESS

The design process consisted of two phases. The first phase was to incorporate the expert recommendations for the redesign of the flight deck and present them in the form of sketch. This phase took the major part of the summer work. The second phase consisted of taking only the space saving ideas for the Forward flight deck from the first phase and implementing and incorporating them in an existing 3D drawing of flight deck.

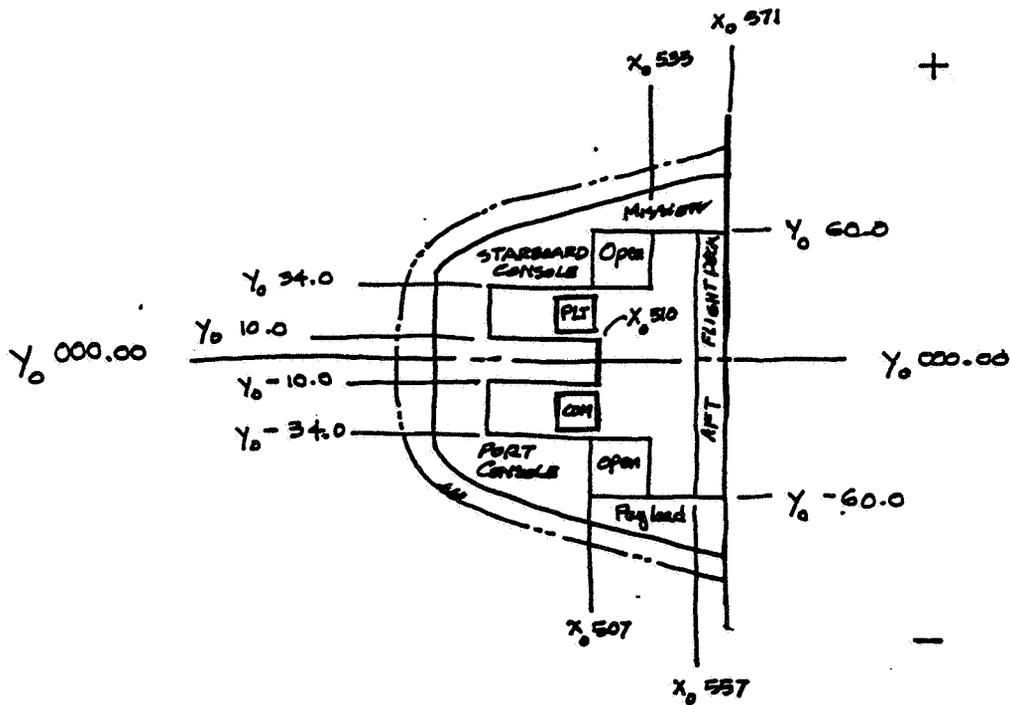
First Phase of Design

The first phase needed much of the data gathering and data organization, physical measurement of the mockup, and study of the original drawings of the flight deck to clearly understand the space we were to work with and design. These drawings were prepared nearly thirty years ago by Rockwell and they are currently used on an every day basis for any alteration or maintenance to the Space Shuttle. The last part of this phase included sketching of the redesign ideas.

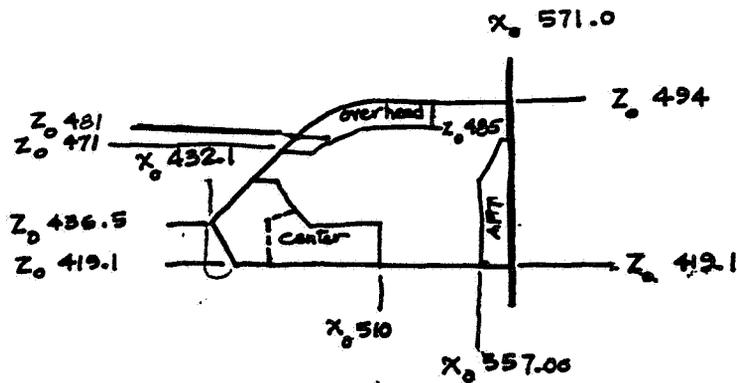
Data Gathering- Several references in form of NASA memos were provided, along with the problem statement. These were studied along with other relevant books, articles, and memos. Although the acquisition of data and documents was sometimes difficult and time consuming, partly due to sensitive nature of some of them, the overall process was a beneficial learning experience. This was an essential stage in the design process for becoming familiar with the space and area one has to work with. It also helped to set the stage for the next step of design which was the actual mockup measurement.

Study of the Original Drawings of the Flight Deck- In order to understand the working space (space to be redesigned) within the flight deck the original drawings of the flight deck by Rockwell were obtained. These drawing were specially useful in providing information about the hidden spaces, which are commonly located under the panels that are to be replaced with more advanced equipment. However, the fact that some of the drawing are very hard to locate, their sizes are not practical for this kind of overall redesign. In order to read original flight deck's drawing one must become familiar with zero base origin for X, Y, and Z coordinates used throughout the Rockwell drawings for flight deck, and at all time take those into account (See Figure 1). Still the exact or even the approximate sizes of the working space (space to redesign) within flight deck could not be easily obtained from the original drawings. While dimensions for some of the individual pieces were given on separate drawings but the overall drawings (Isometrics, sections, and complete views of flight deck such as Forward and Aft) lacked dimensions and specifications. Therefore, after the study of these drawings it was concluded that it would be more efficient to do an actual measurement of the mockup as describe below.

Physical Measurement of the Mockup- The physical measurement of the mockup and identification of its panels started by visiting the building (building 9B) where most mockups are located and video taping the flight deck, mid deck and other relevant areas. The video tape was viewed in order to determine the significant areas to be measured. The actual measurements and part identification of the flight deck took place for a few days on regular basis with the help of some of the engineers on sight. In addition to resolving some of the obscure points in the original drawings, physical measurement of the mockup was very useful in gaining better understanding and becoming more familiar with the redesign of the spaces under study.



Baseline Floor Plan of the Flight Deck



Baseline Section of the Flight Deck

Figure 1. Base Line Drawings (X, Y, and Z Coordinates) of the Flight Deck and Their Origin (Measurements Reference), Rockwell Drawings.

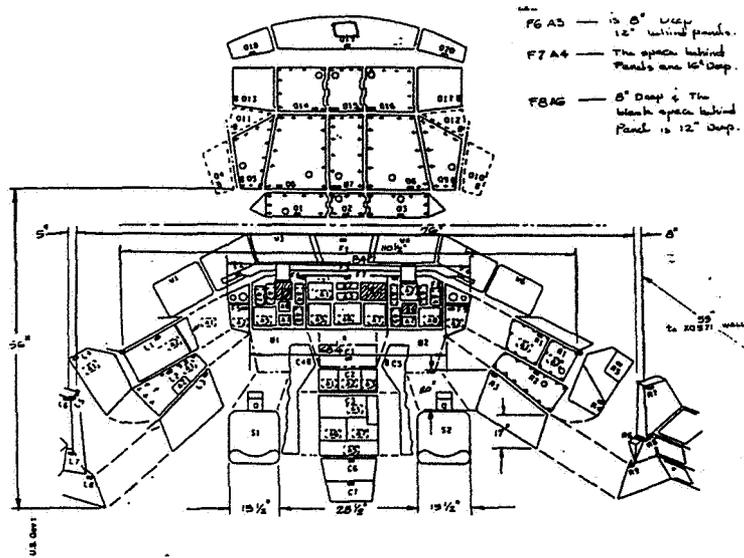
Figures 2 through 5 show the results of physical measurements of the mockup. Each figure is followed by an identical figure on which the location of permanent panels are identified and labeled. Unmarked panels are mission-specific. Some of the permanent panels are subject to future replacement by more advanced equipment saving weight and space and, as a result, cost.

Redesign Recommendations- The following recommendations along with quick sketches were made at the end of this phase of the design.

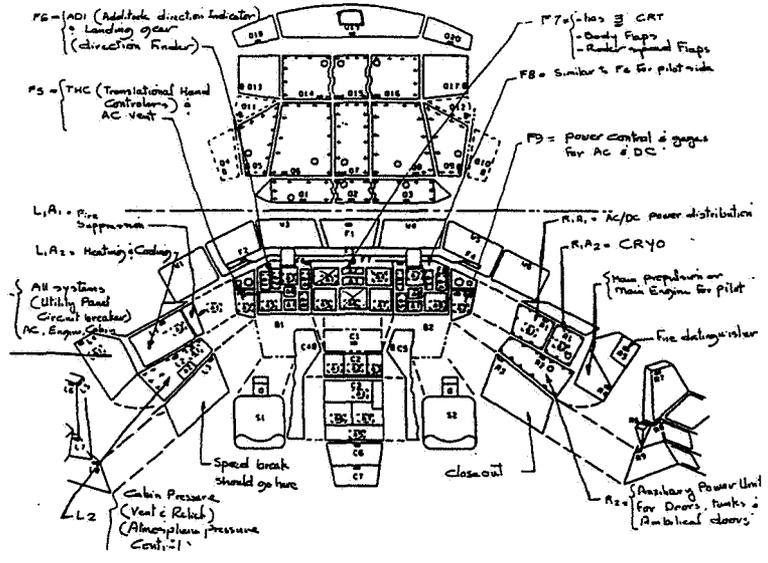
- 1) Flattening the Forward flight deck. The use of flat display panels to replace the existing bulky CRTs. The current dept of the Forward flight deck would be modified to a less space taking panels. This gained space could allow for the pilot and commander seats to be moved further to the front. (See Figure 6)
- 2) Flattening of the AFT flight deck could be another way to save more usable space because the flight deck is at its highest around the AFT station. (See Figure 6)
- 3) Removal of the Right and Left Console could provide a better movement space for the pilot and the commander while seating but not much space is gained in the height direction. (See Figure 6)
- 4) Removal of the center console could provide for one additional seat in front for the mission specialist but it really would take away from the easy communication between the commander and the pilot. It can be heightened and converted to a separate work space for mission specialist. (See Figure 6)
- 5) Removal of the overhead console could be the most useful height-saving idea. It allows for gain in height and opens up the overhead space for more scape panels. (See Figure 6)
- 6) Adding escape panels to the overhead area right above the pilot and the commander seat (converting all seats to ejectable and repositionable by 90 degree seats). These panels would function just like the current overhead window providing additional height. (See Figure 6)
- 7) Design and develop new space saving ejectable seats that can be folded in the floor of the flight deck. This idea requires raising of the flooring of the flight deck which becomes possible after the height gained with the creation of the overhead escape openings. The ejectable seat when not in use should fold in the floor. This would provide for more space for specific missions that need them and for easier movement of the crew members.

Second Phase of Design

The second phase of design concentrated on the space saving ideas for the Forward flight deck (flattening it), Right and Left Console (removing them), and Overhead Console (removing it). (See Figures 7). Most of the redesign ideas in this phase are short-term modifications that do not require any changes to the main body of the flight deck. Therefore, they are interior space saving ideas. These ideas are mostly taken from the previous phase and were given scale and coordinates to

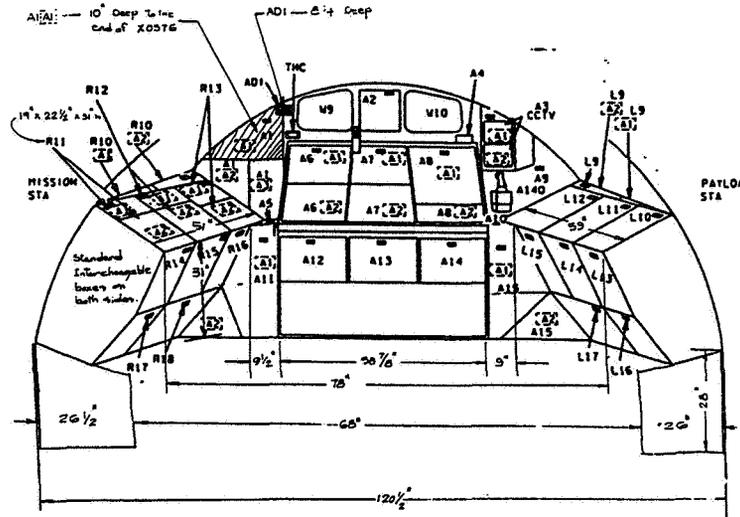


Flight Station Forward Console Dimensions

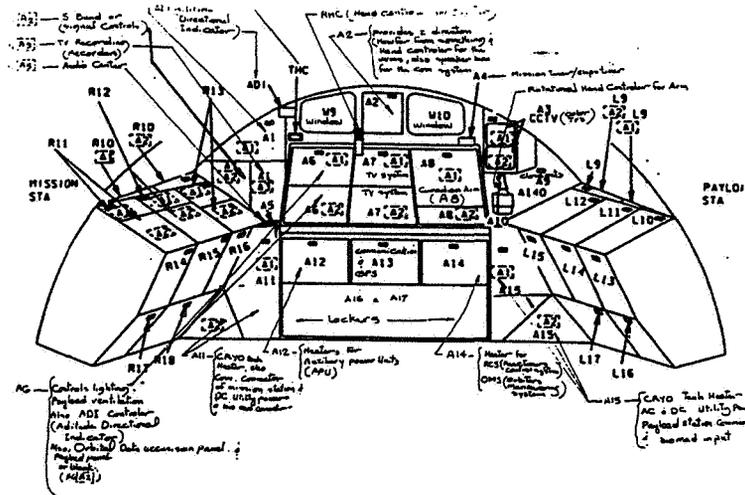


Flight Deck Forward Console Panel Identification

Figure 2. Flight Station Forward Console's Dimensions and Panel Identification.

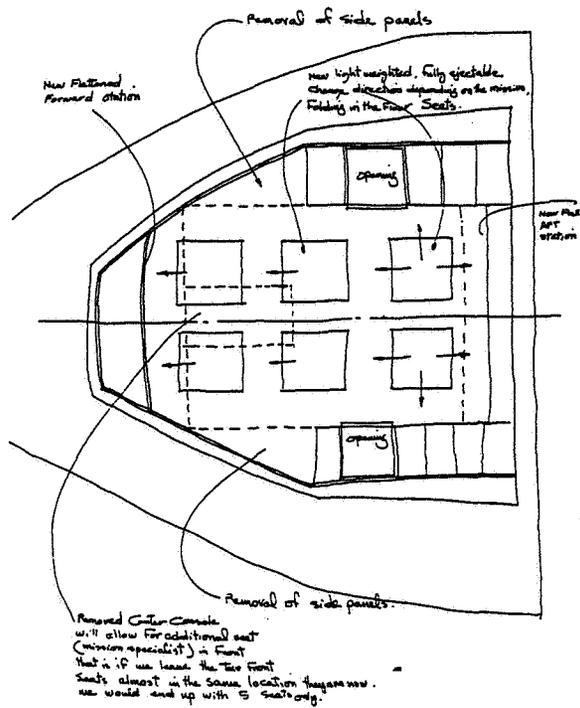


Flight Station AFT Station's Dimensions

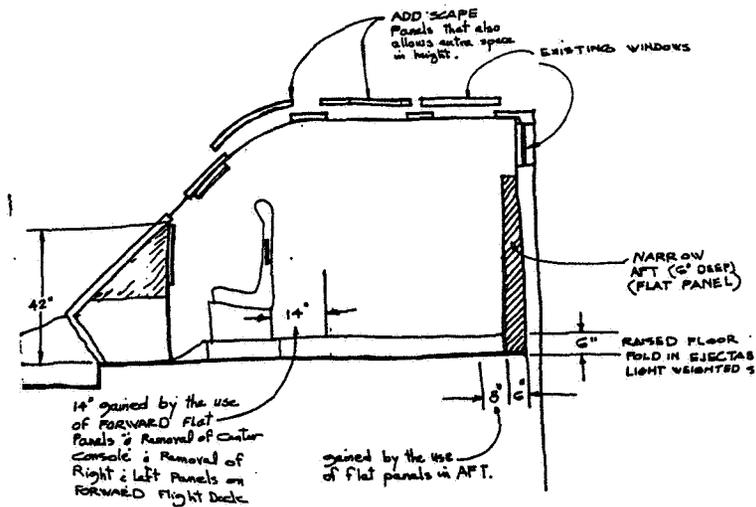


Flight Deck AFT Station Panel Identification

Figure 4. Flight Deck AFT Station's Dimensions and Panel Identification.

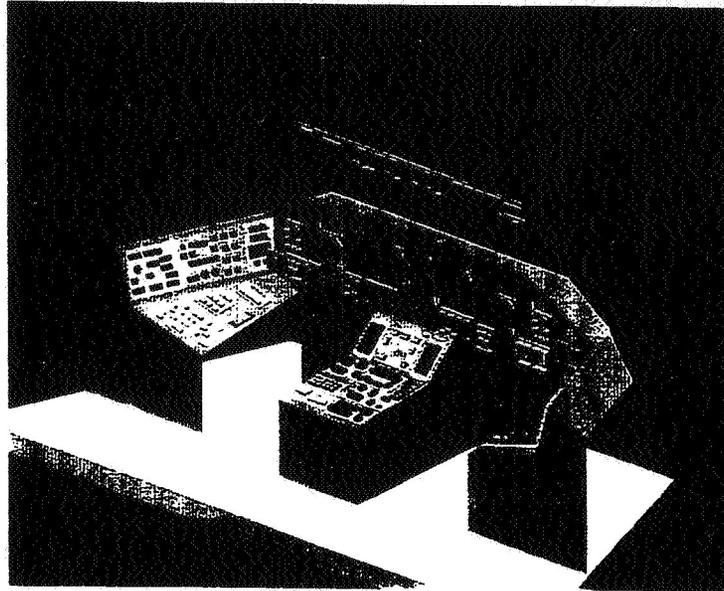


Flight Deck Floor Plan (Redesigned Arrangement)

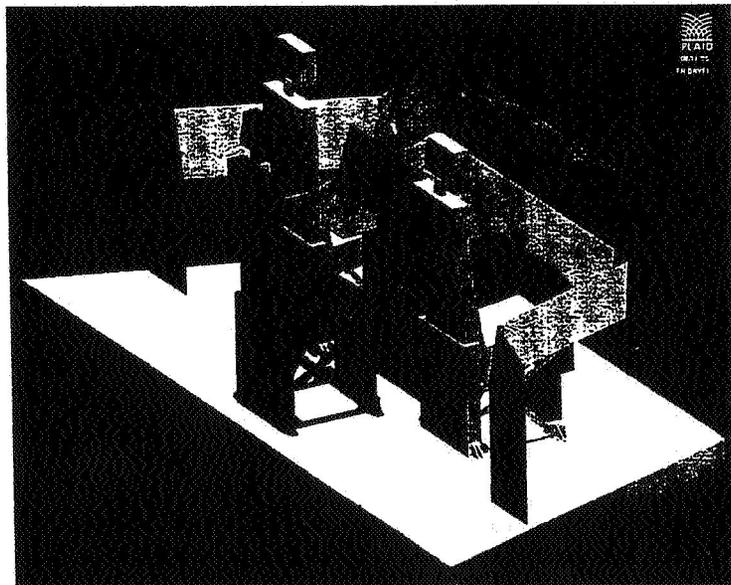


Flight Deck in Section (Redesigned Arrangement)

Figure 6. Flight Deck's Proposed Space Saving Floor Plan Arrangement.



Flight Deck Existing Arrangement



Flight Deck Proposed New Arrangement

Figure 7. Flight Deck Forward, Side Consoles, Overhead, and Center Console Arrangements.

be incorporated on a 3D existing computer drawing of the flight deck. The existing 3D computer drawing were originally created through the use of an in-house developed software called PLAID. This step was done with assistance from one of the GRAPH Lab staff. After modification of the existing drawings based on the short-term redesign ideas, the revised 3D drawing was viewed on Virtual Environment (VE). VE could be a helpful tool in getting the feel of the newly created space without having to build a physical model. It can also be used for future modification by a virtual experience of the actual spaces. The full potential of this tool is yet to be realized.

CONCLUSIONS AND FUTURE RECOMMENDATIONS

Redesign of the Space Shuttle flight deck is a process that requires coming together of many different components. In this project we focused on defining and testing some steps for implementation of this task with particular attention to space-saving concepts. Expert recommendation for overall redesign of Space Shuttle were reviewed and relevant information were extracted and delineated. Original drawings of the Space Shuttle made by Rockwell were obtained and carefully examined. In order to implement and assess any modifications in terms of space saving parameters, it was determined that the drawings alone could not achieve this objective. As a complement, physical measurements of the mockup of Space Shuttle flight deck were made and the information was categorized and properly labeled on the original drawings. Then, space-saving redesign ideas, as motivated by expert recommendations on such things as information display panel upgrade by technologically advanced flat display units, were implemented. Next, the redesign ideas were executed on the forward flight deck, overhead console, right and left console, and center console. A new 3-D computer drawing of this was developed by modifying the existing drawing on the in-house developed software (PLAID). Finally, the drawing was transported to a Virtual Environment and observed.

In order to perform the complete process of redesigning the flight deck exactly, a complete detailed electronic 3-D computer model of the entire current base-line Shuttle flight deck (FWD and AFT) is needed. Such a model could be used for the redesign and further analysis of the Space Shuttle Flight Deck. It should have a data base that would allow the user to be able to add or remove or attach any part(s) within the 3-D model and accurately assess the amount of space (e.g., in cubic feet) and weight (e.g., in pounds) gained or lost. Then redesign idea could be applied much more easily and the changes would be estimated more precisely.

REFERENCES

- (1) "Advanced Trends in Avionics," Jim Curran, Iowa State University Press, 1992.
- (2) Shuttle Avionics Upgrade Investigations, date 7/26/93, Naval Air Warfare Center, Aircraft Division, Indianapolis, IN.
- (3) Clementine: "An Experiment to Flight Quality Lightweight Space Technologies," Pedro L. Rustan.
- (4) JSC memo ACS-90-49, date 11/28/90, "Orbiter Advance Displays".
- (5) JSC memo ACS-94-02, dated 1/10/94, "What Is Obsolete Around Here?"
- (6) JSC memo ACS-93-28, dated 9/21/93, "What Liquid Oxygen/Hydrogen Explosion Tests Are Telling Us"
- (6) "Access to Space-Option 1," Final Presentation to Arnold Aldrich, 8/24/93.

FLUID DYNAMIC VERIFICATION EXPERIMENTS ON STS-70

Final Report

NASA/ASEE Summer Faculty Fellowship Program- 1995

Johnson Space Center

Prepared By:	Stanley J. Kleis, Ph.D.
Academic Rank:	Associate Professor
University and Department:	University of Houston Department of Mechanical Engineering Houston, TX 77204-4792
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Medical Sciences
Branch:	Biomedical Operations and Research
JSC Colleague:	Neal R. Pellis, Ph.D.
Date Submitted:	August 9, 1995
Contract Number:	NGT-44-001-800

ABSTRACT

Fluid dynamic experiments were flown on STS-70 as phase two of the engineering evaluation of the first bioreactor Engineering Development Unit (EDU#1). The phase one experiments were comparative cell cultures in identical units on earth and onboard STS-70. In phase two, two types of fluid dynamic experiments were performed. Qualitative comparisons of the basic flow patterns were evaluated with the use of "dye" streaklines formed from alternate injections of either a mild acid or base solution into the external flow loop that was then perfused into the vessel. The presence of Bromothymol Blue in the fluid then caused color changes from yellow to blue or vice versa, indicating the basic fluid motions. This reversible change could be repeated as desired. In the absence of significant density differences in the fluid, the flow patterns in space should be the same as on earth. Video tape records of the flow patterns for a wide range of operating conditions were obtained.

The second type of fluid dynamic experiment was the quantitative evaluation of the trajectories of solid beads of various densities and sizes. The beads were introduced into the vessel and the paths recorded on video tape, with the vessel operated at various rotation rates and flow perfusion rates. Because of space limitations, the video camera was placed as close as possible to the vessel, resulting in significant optical distortion.

This report describes the analysis methods to obtain comparisons between the in-flight fluid dynamics and numerical models of the flow field. The methods include optical corrections to the video images and calculation of the bead trajectories for given operating conditions and initial bead locations.

INTRODUCTION

The bioreactor development team at NASA/JSC is responsible for the development of a complete cell cultivation system capable of growing and maintaining anchorage dependent cells in a microgravity environment for extended periods of time. The bioreactor system provides control of many parameters required for the successful cell culture while suspending the cells in a fluid environment that allows three dimensional assembly. The present report will address only the fluid dynamics within the culture vessel.

Space flight experiments in STS-70 were scheduled for June, 1995, for the bioreactor developed at NASA/JSC. Unfortunately, the flight was delayed by schedule conflicts with STS-71 and technical problems due to woodpecker damage of the external tank insulation. Two sets of experiments were scheduled; cell growth experiments and fluid dynamic verification experiments.

The current bioreactor vessel design is based in part on the viscous pump reactor vessel developed jointly by NASA/JSC and Dr. S. Kleis of the Turbulent Shear Flow Laboratory (TSFL), University of Houston [1]. The basic elements of the vessel are shown in Figure 1. A three dimensional flow field is established by rotating the outer cylinder and spin filter at different rates. Fluid enters the vessel from the external flow loop, in the gap between the left vessel end and the disc. It then circulates within the vessel before being extracted through the porous spin filter.

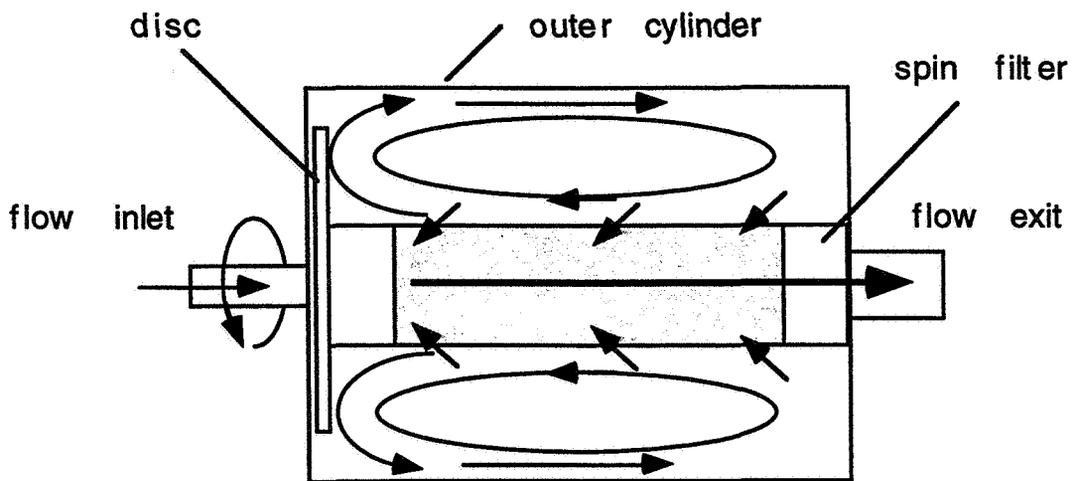


Figure 1.- EDU#1 Bioreactor Flow Fields Elements.

As part of the development of the current vessel design, a numerical model of the flow field within the vessel has been developed. The model has previously been verified under a wide range of operating conditions in a unit gravity environment by extensive measurements of the velocity fields and flow visualizations at TSFL. The purpose of the fluid dynamics phase of the flight experiments was to verify the model under microgravity conditions.

An accurate model of the fluid flow field is required to be able to predict mass transport within the vessel to be able to separate effects of changes in cell hydrodynamic environment from microgravity effects on cells. In the presence of body forces, density differences between the cells attached to micro carriers and the fluid medium cause relative motions, resulting in both mechanical shear and increased mass transport. In the microgravity environment, buoyancy effects are greatly reduced; the normal earth gravity is replaced by centripetal acceleration as the dominant body force. For a typical rotation rate of 2 rev/min in a 5 cm diameter vessel, the magnitude of the body force is reduced to approximately 0.001 m/s^2 compared with gravitational acceleration of 10 m/s^2 on earth. In the absence of other factors, cells would go from a convection dominated mass transport regime on earth to a diffusion dominated regime in microgravity. However, the viscous pump bioreactor vessel has been designed to provide a steady three dimensional flow field with controllable rates of shear. This allows the establishment of local velocity gradients. The local shear flow about the cells can provide control over the mass transfer rates. It is expected that, for most cell types, the shear rates required for adequate mass transport is well below the shear rates that causes damage to the cells by mechanical stress. In fact, it is expected that the shear rates for good mass transport are much lower than the stresses due to cells on micro carriers falling at terminal velocity in earth's gravity. If these characteristics are demonstrated, the bioreactor can be used to study the effects of controlled stress levels on cell function as well as a low stress environment for studies of direct gravity effects on cells.

The objectives of the fluid dynamic experiments conducted on STS-70 are two fold. First, flow visualizations of flow fields established under a wide range of operating conditions will be compared with similar tests previously performed at TSFL. Since

these tests are performed with nearly uniform density, the results can be compared directly with ground based experiments. The second set of fluid dynamic experiments records the positions of beads with diameters of from 1 to 3 mm moving within the bioreactor vessel. The trajectories of these particles were video taped for several different inner and outer wall rotation rates and for several perfusion rates. Unlike the flow visualization tests, these experiments will produce results that can not be obtained on earth. The results of these experiments will also be compared with the numerical model. Differences in the predicted and measured results will provide guidance for modifications to the numerical model.

OPTICAL DISTORTIONS

Conducting experiments on the mid-deck of a shuttle imposes additional constraints on the experiment design. Weight and volume limitations require careful consideration of available supplies and often change procedures. For the fluid dynamic experiments, space considerations meant mounting the video camera as close as possible to the vessel to minimize interference with the crew and other hardware (such as the main hatch, located just to the right of the EDU locker). This required the use of a 3X macro lens, which introduced additional optical distortion (the 'fish eye' effect) into the images. In addition, the bioreactor vessel itself is made of a clear polycarbonate and is filled with a cell culture medium with a refractive index near that of water. The refraction of light by the vessel acts as a cylindrical lens, distorting the apparent positions of cells and beads in the vessel. This effect was greatly reduced by placing an optical correction lens made of polycarbonate between the vessel and camera. The lens had a cylindrical surface slightly larger than the vessel on one side and a flat surface on the camera side as shown in figure 2.

The macro lens distortion can be removed from the position data if the type of distortion is known. If the lens is a simple lens, then the distortion can be approximated as a simple spherical distortion resulting from objects being different distances from the center of the lens.

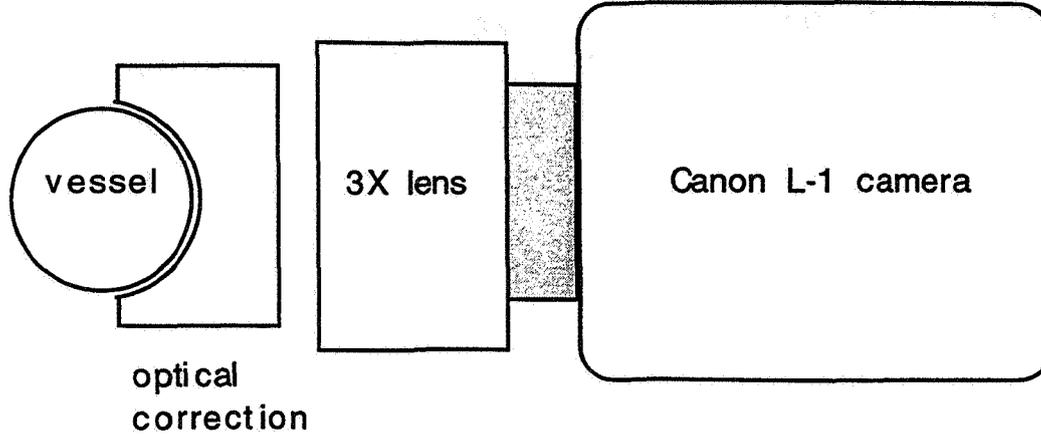


Figure 2.- Optical configure for EDU#1 videos.

For a simple 'thin' lens, the object plane distance O and the image plane distance I , are related to the lens focal length F by the relation,

$$\frac{1}{F} = \frac{1}{O} + \frac{1}{I}.$$

The height of an object in the image plane is the related to the height in the object plane by a similar relation, since

$$\frac{y_o}{O} \propto \frac{y_i}{I}.$$

Assuming that points on a spherical object plane are imaged on a flat image plane, a point at a location (x, y) on a flat object a perpendicular distance O from the lens would have a position (x', y') in the image plane, given by,

$$x' = \frac{xI}{\sqrt{O^2 + x^2 + y^2}}$$

and

$$y' = \frac{yI}{\sqrt{O^2 + x^2 + y^2}}.$$

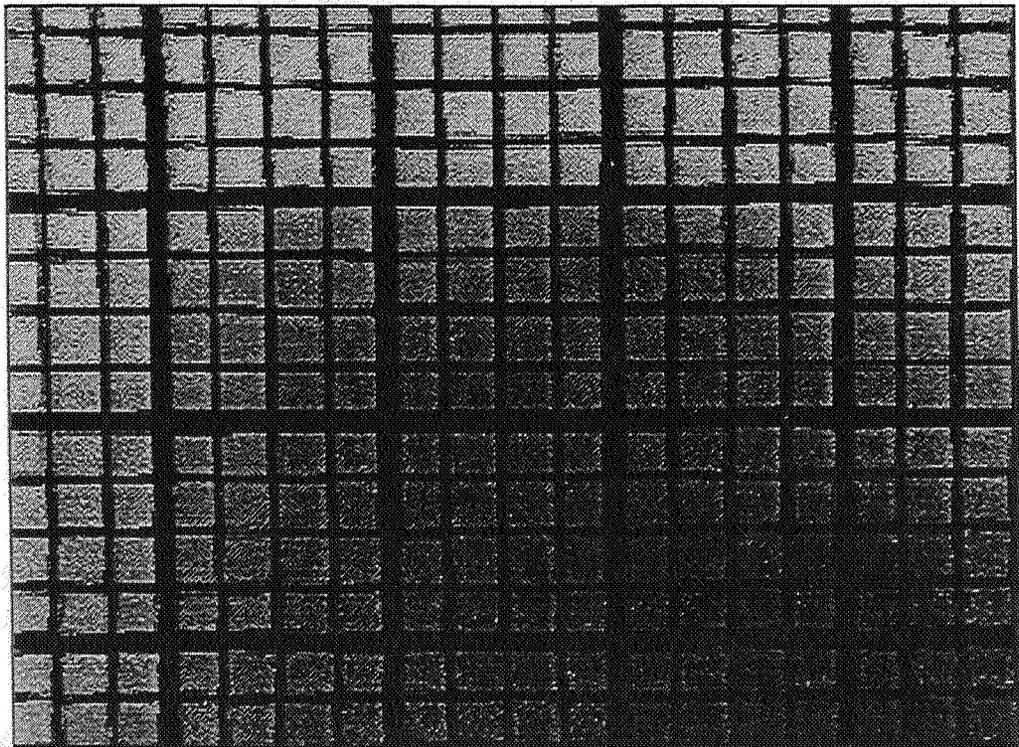
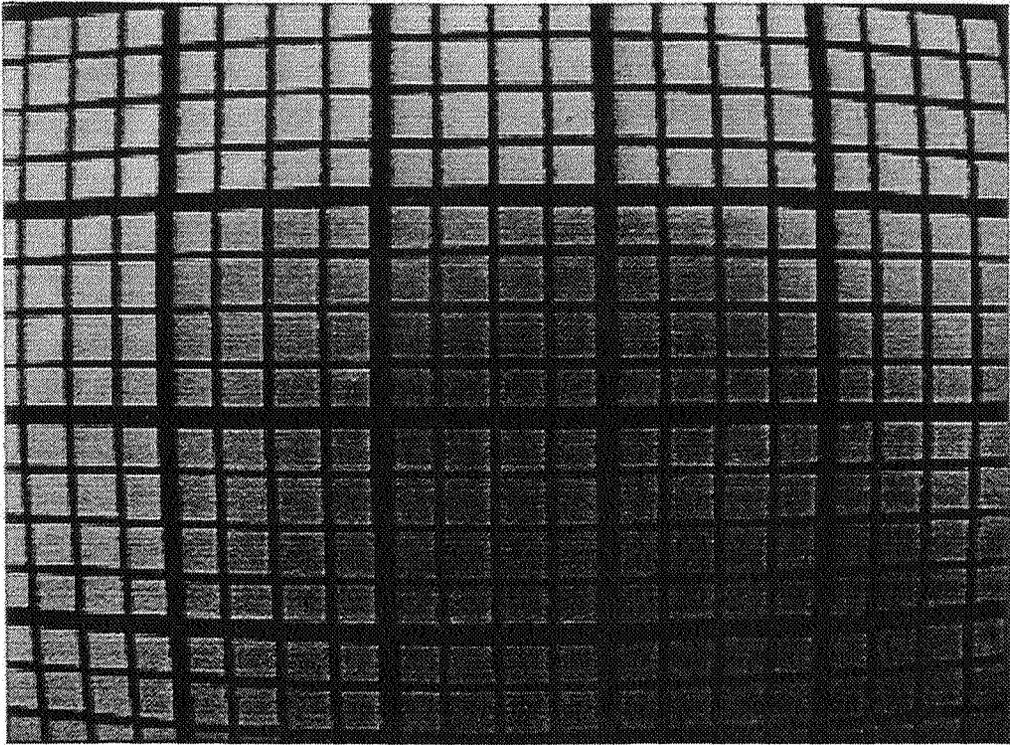


Figure 4.- Grid images before and after correction.

To see if this simple spherical correction would be adequate for the present study, a regular grid with lines at 0.5 in was placed in front of the flight camera at the same distance and camera settings to be used for the actual data recording. The digitized image before and after correction is shown in Figure 4. It appears that the spherical correction does a good job of removing the lens induced distortions. However, applying the correction to particle position results will require knowledge (or estimates) of the out-of-plane position of the beads. For the present study, the actual corrections will be applied to the measured positions from the original (distorted) images using estimates of the out-of-plane position of the beads. Thus, the position data will be corrected, not the images.

PARTICLE POSITION ESTIMATES

As discussed above, the relative motion between the fluid and the cells on microcarriers can have a dramatic effect on the amount of mass transport to and from the cells. In the absence of gravity or other body forces to drive convection currents, mass transport in a fluid without shear relies on diffusion. A solid sphere in a viscous flow field with shear will have a region of fluid that surrounds and moves with the sphere (a closed stream surface). When shear is present, the distance over which diffusion takes place is reduced and concentration gradients that drive the diffusion are increased. In a rotating vessel, the shear and relative motion caused by density differences in the centripetal acceleration field will determine the mass transport. Since the centripetal acceleration is about 10^4 smaller than earth's gravity, the mechanical stress on cells can be reduced to very low levels compared with those associated with the cells on microcarriers falling through the fluid in a 1 g field, by lowering the rotation rates of the spin filter and outer cylinder (see Figure 1).

The bead trajectory studies will give quantitative results that can be compared with predictions from computer models. Bead positions are estimated from force balances of the beads in the computed velocity fields for the prescribed operating conditions.

The general force balance equation for a sphere in an unsteady, non uniform flow field is [2],

$$m_p \frac{dV_i}{dt} = m_f \frac{Du_i}{Dt} - \frac{1}{2} m_f \frac{d}{dt} \left\{ V_i - u_i - \frac{1}{10} a^2 \nabla^2 u_i \right\} - 6\pi a \mu \{ V_i - u_i \} + \pi a^3 \mu \nabla^2 u_i$$

Where m_p is the mass of the particle, m_f is the mass of fluid displaced by the sphere, a is the sphere radius, V_i is the i th sphere velocity component, u_i is the corresponding fluid velocity component evaluated at the current position and time, and μ is the fluid viscosity. Note: The Basset history integral term, which accounts for the transient decay of the initial conditions, has been neglected in this equation.

When performing a force balance in cylindrical coordinates, the proper equations relating the net force to time rates of change of position are:

$$F_r = m \left(\frac{d^2 r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \right), F_\theta = m \left(r \frac{d^2 \theta}{dt^2} + 2 \frac{dr}{dt} \frac{d\theta}{dt} \right), \text{ and } F_z = m \frac{d^2 z}{dt^2}.$$

Normalizing, the resulting component equations are:

$$\left(1 + \frac{1}{2} \frac{\rho_f}{\rho_s} \right) \left(\frac{d^2 r}{dt^2} - r \left(\frac{d\theta}{dt} \right)^2 \right) = \frac{\rho_f}{\rho_s} \frac{Du_r}{Dt} + \frac{1}{2} \frac{\rho_f}{\rho_s} \frac{d}{dt} (u_r) + \frac{a^2}{20 r^2} \frac{\rho_f}{\rho_s} \frac{d}{dt} (\nabla^2 u_r) - \frac{9\mu r_2}{2U\rho_s a^2} (v_r - u_r) + \frac{3\mu}{4U\rho_s r_2} (\nabla^2 u_r)$$

$$\left(1 + \frac{1}{2} \frac{\rho_f}{\rho_s} \right) \left(r \frac{d^2 \theta}{dt^2} + 2 \frac{dr}{dt} \frac{d\theta}{dt} \right) = \frac{\rho_f}{\rho_s} \frac{Du_\theta}{Dt} + \frac{1}{2} \frac{\rho_f}{\rho_s} \frac{d}{dt} (u_\theta) + \frac{a^2}{20 r^2} \frac{\rho_f}{\rho_s} \frac{d}{dt} (\nabla^2 u_\theta) - \frac{9\mu r_2}{2U\rho_s a^2} (v_\theta - u_\theta) + \frac{3\mu}{4U\rho_s r_2} (\nabla^2 u_\theta)$$

and

$$\left(1 + \frac{1}{2} \frac{\rho_f}{\rho_s} \right) \frac{d^2 z}{dt^2} = \frac{\rho_f}{\rho_s} \frac{Du_z}{Dt} + \frac{1}{2} \frac{\rho_f}{\rho_s} \frac{d}{dt} (u_z) + \frac{a^2}{20 r^2} \frac{\rho_f}{\rho_s} \frac{d}{dt} (\nabla^2 u_z) - \frac{9\mu r_2}{2U\rho_s a^2} (v_z - u_z) + \frac{3\mu}{4U\rho_s r_2} (\nabla^2 u_z)$$

$R_0=2$ cm, $Z_0=2.2$ cm

$R_0=2.2$ cm, $Z_0=3$ cm

$R_0=2.3$ cm, $Z_0=3$ cm

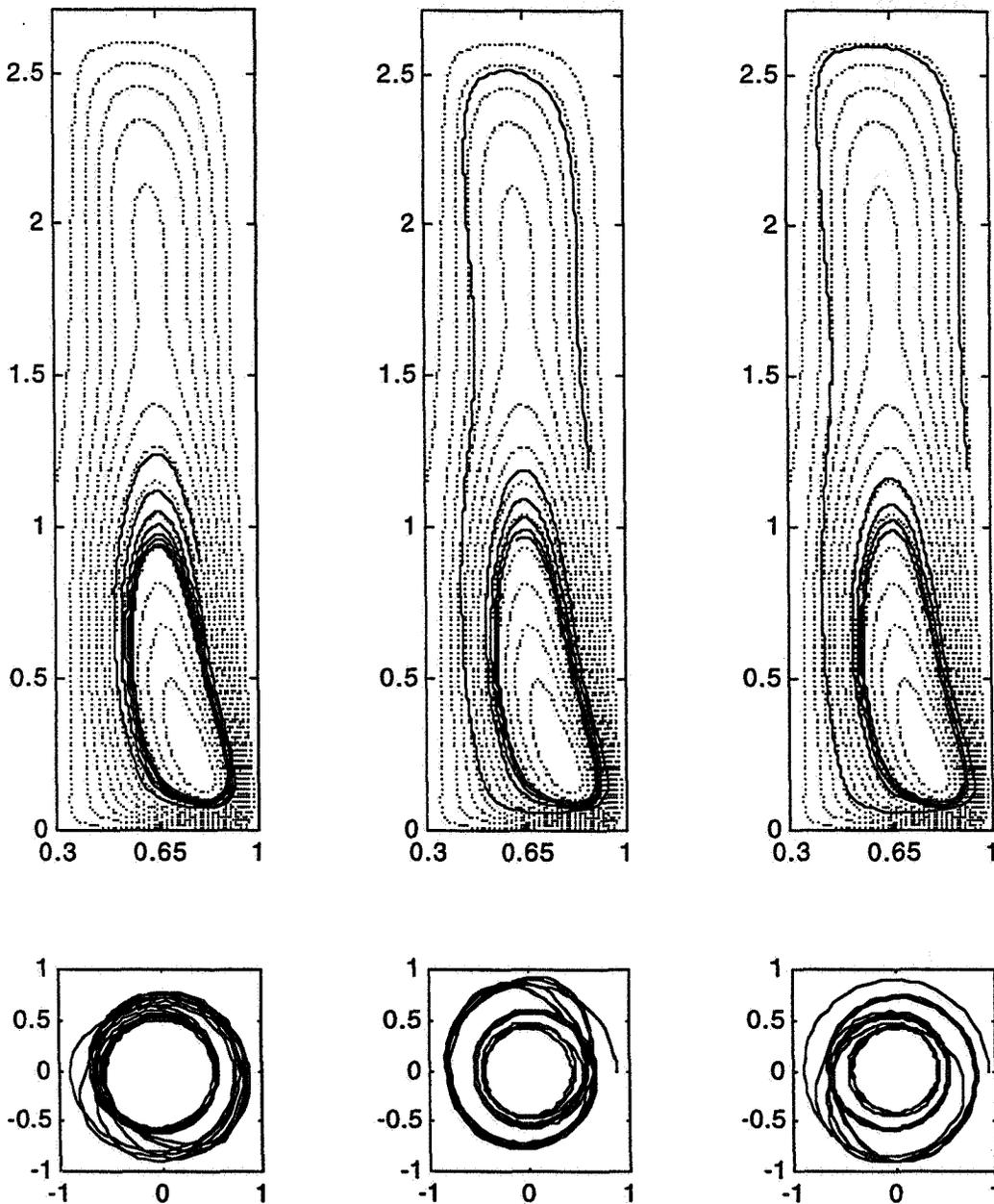


Figure 5.- Particle Trajectories for three initial particle positions (shown above figures) for the case of disc rotation rate = 11 rpm, outer cylinder rotation rate = 0.5 rpm, perfusion = 0 cc/min, bead diameter = 3.175 mm, $\rho_s/\rho_f = 1.045$. Total time = 5 min.

$R_o=2$ cm, $Z_o=2$ cm $R_o=2.3$ cm, $Z_o=3$ cm $R_o=1.875$ cm, $Z_o=1.875$ cm

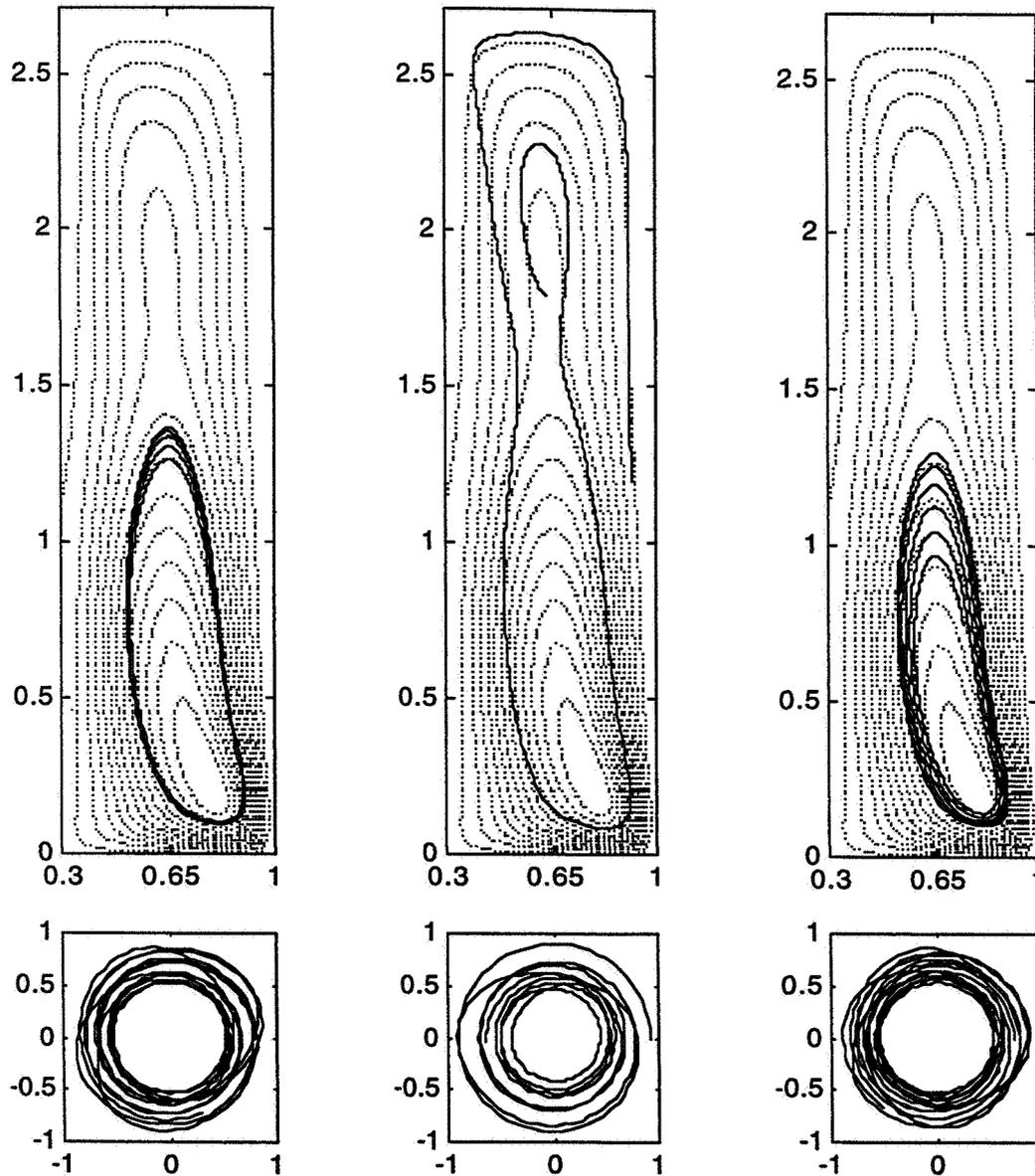


Figure 6.- Heavier particle trajectories for three initial particle positions (shown above figures) for the case of disc rotation rate = 11 rpm, outer cylinder rotation rate = 0.5 rpm, perfusion = 0 cc/min, bead diameter = 3.175 mm, $\rho_s/\rho_f = 1.18$. Total time = 5 min.

$R_0=2$ cm, $Z_0=3.75$ cm $R_0=1.8$ cm, $Z_0=1.5$ cm $R_0=2$ cm, $Z_0=2$ cm

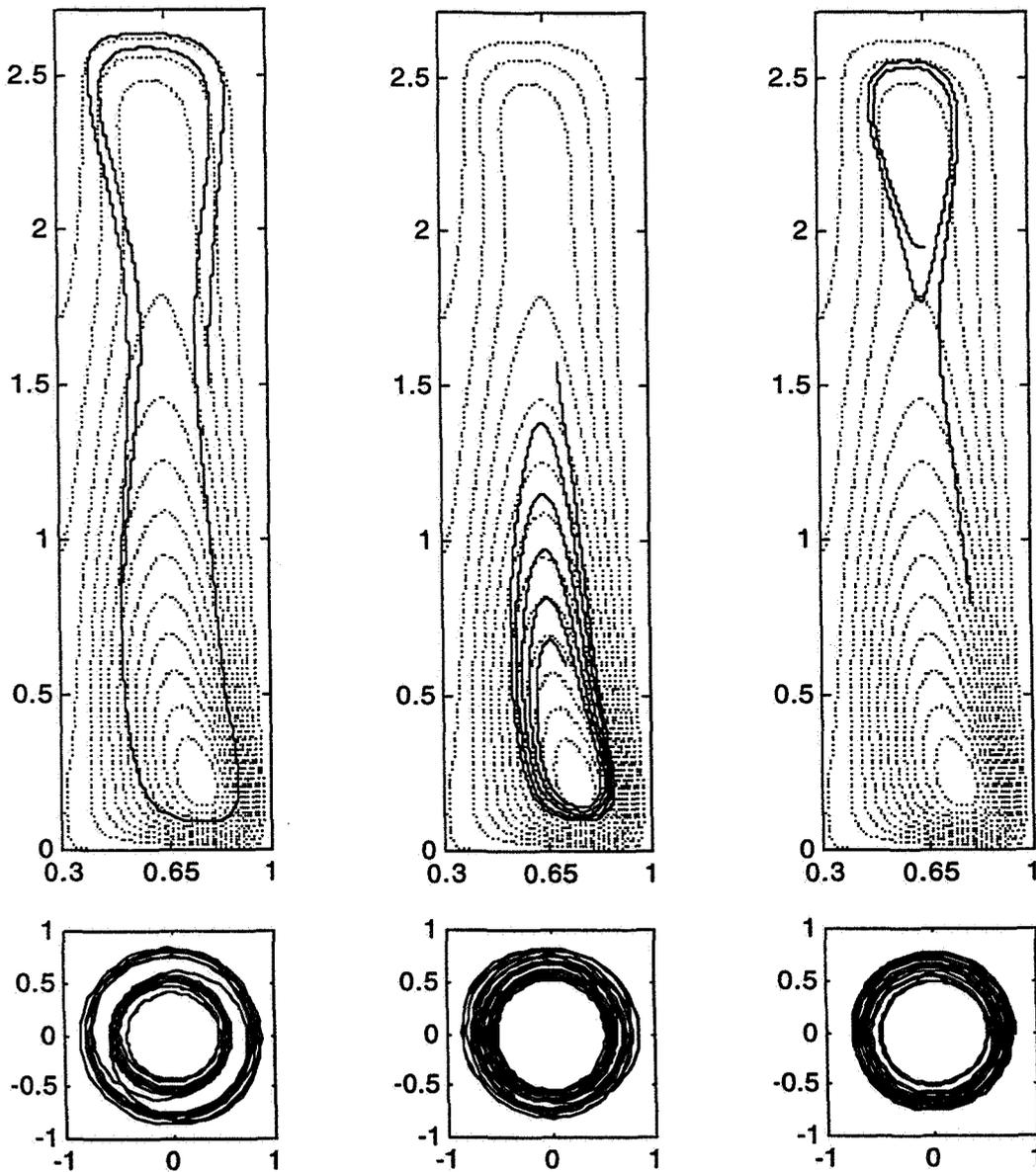


Figure 7.- Heavier particle trajectories for three initial particle positions for the case of disc rotation rate = 6 rpm, outer cylinder rotation rate = 1.0 rpm, perfusion = 10 cc/min, bead diameter = 3.175 mm, $\rho_s/\rho_f = 1.18$. Total time = 10 min.

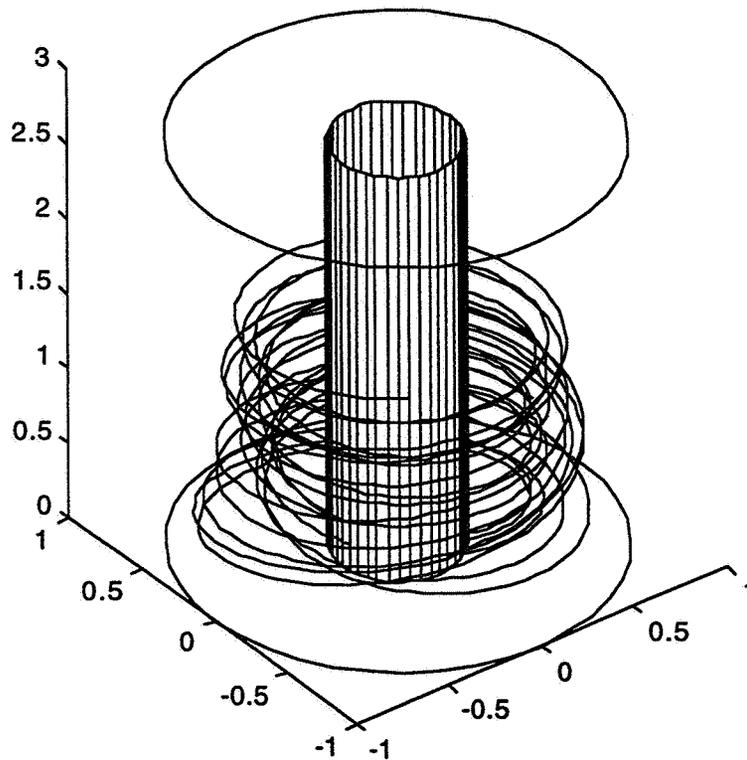


Figure 8.- Heavier particle trajectory for initial particle position (1.8 cm, 0, 1.5 cm) for the case of disc rotation rate = 6 rpm, outer cylinder rotation rate = 1.0 rpm, perfusion = 10 cc/min, bead diameter = 3.175 mm, $\rho_s/\rho_f = 1.18$. Total time = 10 min.

In these equations the fluid and sphere velocities have been normalized by the characteristic velocity, U , the larger of the tip velocity of the disc or outer cylinder, and the length scale used was r_2 , the radius of the outer cylinder of the bioreactor (2.5 cm for the current design). The subscripts s and f refer to the sphere or fluid, respectively.

These equations were solved using a fourth order Runge-Kutta method for the position coordinates as functions of time using the fluid velocities computed by a previous numerical model for steady fluid motion. Figures 5, 6, and 7 show the bead trajectories in side

views, (r, z) planes, in the top figures and top views, (r, θ) planes, in the lower figures. In each figure, three initial bead positions are shown. Figure 5 represents the conditions used for the cell experiments on STS-70, with a bead density ratio near that of cells on microcarriers. Figure 6 is for the same operating conditions, but for a bead whose density ratio and size match the bead used in the fluid dynamic experiments of STS-70. Figure 7 is for a lower rotation rate, typical of future cell studies. In all cases, the bead is suspended in the fluid (stays off the outer wall) with significant axial and radial motions. Figure 8 shows a perspective view of the three dimensional bead motion.

It is interesting to note that the equations of motion show that neutrally buoyant larger beads in a general three dimensional flow field will not follow the streamlines. The terms including the Laplacian of the fluid velocity do not approach zero as the bead density approaches that of the fluid. Thus, a zero relative velocity is not a solution for neutrally buoyant bead. These terms do approach zero in the limit of small bead size. This is a reflection of the differences between a rigid sphere and a deformable fluid element.

Several test cases have been used to verify the accuracy of the bead position calculations. These include flow fields for solid body rotation and simple Couette flow with infinitely long cylinders. The results agreed as expected. Unfortunately, there is no known exact solution which has all three velocity components. The predictions do, however, behave properly (follow streamlines) for the case of small beads in a computed three dimensional flow.

CONCLUSIONS

It has been shown that a simple spherical optical correction for the macro lens is adequate for processing bead position data. Estimates of the out-of-plane coordinate will be necessary to process the bead position results from STS-70.

Bead position predictions based upon previous numerical models of the flow field and force balances on the beads have been developed for comparison with the flight experiments. The flight experiment video tapes have just become available. However, preliminary results indicate that the predicted trajectories look qualitatively similar to the bead trajectories observed in the down

linked videos from the flight experiments. More quantitative comparisons are being made.

REFERENCES

1. "A Viscous Pump Bioreactor," S. Kleis, S. Schreck, and R. Nerem, Biotechnology and Bioengineering, Vol. 36, pp. 771-777 (1990).
2. "Equation of Motion for a Small Rigid Sphere in a Nonuniform Flow," M. R. Maxey and J. J. Riley, Phys. Fluids, 26, pp. 883-889, (1983).

LCTV HOLOGRAPHIC IMAGING

**Final Report
NASA/ASEE Summer Faculty Fellowship Program - 1995
Johnson Space Center**

Prepared By: Jerome Knopp, Ph.D.

Academic Rank: Associate Professor

**University & Department: The University of Missouri-Columbia/Kansas City
CEP Electrical and Computer Engineering Dept.
Kansas City, Missouri 64110**

NASA/JSC

Directorate: Engineering

Division: Avionic Systems

Branch: Electromagnetic Systems and Navigational Sensors

JSC Colleague: Richard D. Juday

Date Submitted: October 26, 1995

Contract Number: NGT-44-001-800

ABSTRACT

Astronauts are required to interface with complex systems that require sophisticated displays to communicate effectively. Lightweight, head-mounted real-time displays that present holographic images for comfortable viewing may be the ideal solution. We describe an implementation of a liquid crystal television (LCTV) as a spatial light modulator (SLM) for the display of holograms. The implementation required the solution of a complex set of problems. These include field calculations, determination of the LCTV-SLM complex transmittance characteristics and a precise knowledge of the signal mapping between the LCTV and framegrabbing board that controls it. Realizing the hologram is further complicated by the coupling that occurs between the phase and amplitude in the LCTV transmittance. A single drive signal (a gray level signal from a framegrabber) determines both amplitude and phase. Since they are not independently controllable (as is true in the ideal SLM) one must deal with the problem of optimizing (in some sense) the hologram based on this constraint. Solutions for the above problems have been found. An algorithm has been for field calculations that uses an efficient outer product formulation. Juday's MEDOF⁷ (Minimum Euclidean Distance Optimal Filter) algorithm used for originally for filter calculations has been successfully adapted to handle metrics appropriate for holography. This has solved the problem of optimizing the hologram to the constraints imposed by coupling. Two laboratory methods have been developed for determining an accurate mapping of framegrabber pixels to LCTV pixels. A friendly software system has been developed that integrates the hologram calculation and realization process using a simple set of instructions. The computer code and all the laboratory measurement techniques determining SLM parameters have been proven with the production of a high quality test image.

LCTV HOLOGRAPHIC IMAGING

Astronauts are required to interface with complex systems that require sophisticated displays to communicate effectively. Display systems that can render 3-D objects offer significant advantages over 2-D displays for certain types of information. Holographic images with true parallax provides a high quality image that can be observed for long periods of time without the discomfort that sometimes occurs in conventional stereograms. Furthermore, a single hologram can replace a complex optical system and can reduce the weight and the number of optical components. Compact lightweight optical systems are the key to building practical head-mounted display systems that are attractive for astronaut use. Holographic displays can simultaneously incorporate the optics and the imaging device in a single element. The ideal holographic display is a high resolution spatial light modulator (SLM) with real-time imaging capability. One particular attractive and economic SLM that has real-time capability is liquid crystal television (LCTV). We describe here an implementation of an LCTV for the display of holograms. This implementation required the solution of a complex set of problems. These include field calculations, determination of the LCTV-SLM complex transmittance characteristics and a precise knowledge of the signal mapping between the LCTV and framegrabbing board that controls it. Realizing the hologram is further complicated by the coupling that occurs between the phase and amplitude of the LCTV. A single drive signal (a gray level signal from the framegrabber) determines both amplitude and phase. Since they are not independently controllable one must deal with the issue of optimizing (in some sense) the hologram based on this constraint. The solutions for these problems is discussed below.

MEASUREMENT OF THE LCTV COUPLING CURVE

Before the LCTV can be used as an SLM the amplitude and phase characteristics of its complex transmittance must be determined by laboratory measurement. The LCTV screen is operated between two polarizers; the angles of the polarizers determine the SLM characteristics. Each combination of polarizer angles determines a unique SLM operating curve. The gamut of operating curves varies considerably and includes highly coupled, phase-mostly and amplitude-mostly curves. But in all cases the phase and amplitude are not independent; they are controlled by a single signal, a gray level value from a framegrabber board. In order to assure high light throughput and a bright image, the polarizers were set for phase-mostly operation. This was not necessarily the best choice of SLM curve in terms of image quality but it did assure that the image would be easy to locate. The polarizer angles for phase-mostly operation were determined by previous experiences with other LCTV cells¹.

SLM Measurements

The determination of the SLM curve was based on separate measurements of the amplitude and phase characteristics as a function of gray level value. The amplitude transmittance was determined from intensity transmittance measurements of a helium-neon laser beam passing through the LCTV. The relative phase transmittance was inferred from fringe shifts measurements made with a grating interferometer. This approach is described in detail elsewhere ¹. The shift measurements were made from framegrabbed images of fringes. In the past, ad-hoc techniques requiring a mix of computer analysis and human intervention were used to determine the shifts by examining null positions in the pattern. Because the patterns are noisy and are not uniform in amplitude variation across the fringe image the null positions can be difficult to define. This required determining null neighborhoods by eye and then using the computer to fine tune the search process. A new approach was tested that avoided the search process.

Algorithms for Fringe and Phase Analysis

A first attempt was made at developing software that carried out the fringe analysis completely without human intervention. The fringe periodicity and the shift were determined using correlation. The correlations produced fairly smooth curves with clearly defined maxima and minima. The fringe pattern with the greatest contrast was used as a base on which to correlate other shifted patterns. The proper scale relation between fringe shift and phase shift was determined by an accurate measurement on the periodicity of fringe patterns. The average periodicity from the autocorrelations of all the fringe patterns gave a good estimate of the period. This was done using the Matlab m-file SMOOTH.M. The shift in the peak (global maxima) of the cross-correlation of all the patterns with the reference fringe pattern determined the relative phase shift. The technique worked well with noisy fringe data with only one caveat. The non-uniformity in the fringe contrast sometimes results in large jumps of nearly one period in the peak position due to small differences in local maxima. This was fixed by linearly interpolating phase shifts where large jumps occurred. The m-file FRINGE4.M calculated the cross-correlations and the peak shifts. The program has only been used on one data set and needs further testing. But it ran without any problems and produced data that appeared consistent with the observed fringe shift measurement estimated by ruler and eye.

CALCULATION OF A REALIZABLE HOLOGRAM

Most of the past work in calculating fields for computer generated holograms is associated with farfield or Fourier transform holograms ^{2,3}. This is primarily due to the low resolution of SLMs available and the low resolution requirements

of Fourier transform holograms. There is not a significant corresponding body of work on nearfield computer generated holograms. The general problem for calculating the nearfield for an arbitrary object source is not too difficult for planar objects where the problem can be posed as a linear filter problem⁵. In this case the solution can be implemented on computer where the efficiencies inherent in the fast Fourier transform can be exploited. For three dimensional objects it is often more efficient to define the object as a collection of elemental objects such as points, lines, rectangular apertures, etc. and use the superposition of elemental fields to find the object field. We have chosen this approach, but have exploited approximations that lead to separability of mathematical operations in the x and y directions. This separability is inherent in field calculations for rectangular apertures and point objects modeled with quadratic approximations. This results in an efficient outer product formulations that are easy to implement in Matlab.

The Outer Product Formulations

Calculations were made for two letter "F" objects. One defined as a collection of rectangular apertures, the other as a set of point sources. The superposed fields for the elemental objects (rectangular apertures or point sources) determines the total field for each F.

Rectangular Apertures

In the case of rectangular objects the fields for the rectangles were estimated using small angle (quadratic) approximations and formulated in terms of separable Fresnel integrals. The phasor field for a single rectangular aperture assumed to be transilluminated by a plane wave (or self-luminous with a uniform field across the aperture) is⁵:

$$U(x_o, y_o) = A \frac{e^{jkz}}{j\lambda z} \int_{-x_{lc}}^{x_{uc}} e^{\frac{jk}{2z}(x_1 - x_o)^2} dx_1 \int_{-y_{lc}}^{y_{uc}} e^{\frac{jk}{2z}(y_1 - y_o)^2} dy_1 \quad (1)$$

Here x_o, y_o are the coordinates of point at which the field is desired, A is a constant, z is the distance to from the observation point from the plane of the rectangular aperture, and λ is the wavelength and k is the propagation constant $2\pi/\lambda$. The limits of the integrals, x_{lc}, y_{lc} , and x_{uc}, y_{uc} , are the coordinate pairs of the lower left corner and the upper right corner of the rectangular aperture. The integrals in Eq. (1) can be redefined in terms of ξ, η with a change of variables defined by the following relations:

$$\frac{\pi \xi^2}{2} = \frac{k}{2z}(x_1 - x_0)^2$$

$$\frac{\pi \eta^2}{2} = \frac{k}{2z}(y_1 - y_0)^2$$

This gives:

$$U(x_0, y_0) = A \frac{e^{jkz}}{2} \int_{-\xi_1}^{\xi_2} e^{j\frac{\pi}{2}\xi^2} d\xi \int_{-\eta_1}^{\eta_2} e^{j\frac{\pi}{2}\eta^2} d\eta \quad (2)$$

The integrals in Eq. (2) are the complex form of the tabulated Fresnel integrals. Pade approximations to the Fresnel integrals are known. We used the approximation by Hastings⁶ in the m-file FRES4.M to estimate the integrals and to calculate the fields. FRES4.M is called by the m-file FRESNEL.M which is called with the upper and lower corner coordinates of an elemental rectangular aperture as arguments. The superposition of the elemental fields (arranged to form a letter F) is given in the m-file F_FIELD.M. The first calculations from this approach produced a field intensity that showed a diffracted and blurred projection of the letter F with fringing at the edges of the letter F. The pattern produced looked very much like the in-line holograms produced originally by Gabor⁴. It was decided that using this particular type of hologram might be a problem as an initial test object. Since one would observe the virtual image of the F looking through a hologram that itself resembled an F, the possibility of confusion between the image in the hologram and the virtual image exists. It was decided to use a diffusive representation of the letter F instead. The diffusive virtual image consisted of a set of randomly phased point sources arranged to form an F. The use of random phasing guaranteed a specular field intensity in the hologram plane in which the underlying structure of the F was completely hidden from the eye.

Point Sources

Holograms resulting from point sources can be efficiently calculated using a quadratic approximation to each point source. The quadratic approximation leads to the same x-y separability that was exploited in the rectangular aperture calculations. Given a point source at of strength A located at (x_1, y_1) the phasor field can be represented U_p is given by⁵:

$$U_p(x_o, y_o) = \frac{A e^{jk[(x_o-x_1)^2+(y_o-y_1)^2(z_o-z_1)^2]}}{\sqrt{(x_o-x_1)^2+(y_o-y_1)^2(z_o-z_1)^2}}$$

If we assume that the point source is located near the z-axis and that the point observation makes small angles with respect to the z-axis then the field may be approximated as ⁵:

$$U(x_o, y_o) = \frac{A}{z} e^{j\frac{k}{2z}(x_o-x_1)^2} e^{j\frac{k}{2z}(y_o-y_1)^2}$$

As in the previous case, the separability of the quadratic approximation implies that an outer product formulation is possible. The Matlab code for these calculations is in the m-file POINTHOL.M.

Optimization of the Hologram: HOLOMED

Once a hologram has been calculated, its field must be optimally fitted to the SLM characteristics of the LCTV. The fitting process was carried out by optimizing the ratio of the light throughput of the hologram to the square error in the Euclidean distance in the complex plane between the desired complex transmittance and the realizable transmittance of the LCTV. This optimization was carried out using an adaptation of the filter program MEDOF (Minimum Euclidean Distance Optimal Filter) developed Juday⁷. His adaptation of MEDOF dubbed HOLOMED will be described in a future paper⁸. The output of the HOLOMED program is a 220(rows) X 320 array of gray level values ranging from 0 to 255. These gray level values must be written to the LCTV via a framegrabber.

THE LCTV - FRAMEGRABBER AFFINE MAP

It would be convenient if the gray level values could be written to the LCTV with a framegrabber that mapped gray level values one to one from framegrabber pixels to LCTV pixels. This is unfortunately not case. The Matrox⁹ PIP framegrabber used for writing to the LCTV requires a 512 X 512 array of gray level values from which a video signal is synthesized for the LCTV. Only a portion of the PIP array values are mapped to the LCTV. Rows in the PIP are converted to a standard video lines that are written to the LCTV. Odd rows and even rows separately determine the two fields that determine the interlaced image in a conventional television frame. The LCTV has only 220

rows of pixels. It overwrites the two fields on top each other; it does not interlace. Therefore 440 rows of PIP pixels are mapped to the 220 rows of the LCTV. Determining which rows in the framegrabber mapped to which lines in the LCTV was easily determined by consecutively writing "lines of pixels" (All pixels in a line set to a gray level of 255, remaining pixels to 0) from the PIP. to the LCTV and observing which PIP row first turned on the top row of pixels on the LCTV. This completely defined the row to row map. It does not define the details of the pixel mapping. Determining how pixels within a given PIP row mapped to pixels within an LCTV row was more difficult. The problem is to determine precisely where in a given line video signal from the PIP starts and stops writing to LCTV row. The start and stop points are arbitrary and there locations depend on the details of the timing circuits in the PIP and the LCTV drive electronics. This requires a measurement with subpixel accuracy. This measurement was done in two steps.

First, a rough measurement was made that resolved the start and stop positions in the framegrabber to within a framegrabber pixel. by consecutively "turning on" (i.e. set one pixel to 255, the rest to 0) single pixels in a framegrabber row. The effect on the LCTV screen pixels were then observed through a television microscope system. A change in the first pixel of the LCTV line. indicated the start position of PIP line writes while a change in the last LCTV pixel on the line determined the stop position for line writes. This produced a rough estimate of the. mapping ratio of PIP pixels to LCTV pixels. Two other methods were then used to fine tune the mapping between framegrabber and LCTV pixels and achieve subpixel resolution.

The first method, developed by Juday⁸, was based on writing a pattern of sinusoids of varying frequencies and phase shifts to the LCTV screen. The patterns were changed until a sinusoid was found that produced a uniform intensity across the LCTV of two pixels on and two pixels off. A slight error in either phase or frequency produced an easily discernable gross patten non-uniformity that was easy to see when the LCTV screen was magnified.

A second method, developed by Knopp⁸, was based on observing a moiré. frequency of a created by a bar pattern with a 6 pixel period. (3 pixels on, 3 pixels off) that beat with between the LCTV screen periodic pixel pattern. A low frequency beat of a few cycles across the LCTV screen was observed due to a low frequency beat in the 3rd harmonic of the periodic pattern with the LCTV pixel pitch. This permitted a fairly good measurement of the ratio of the PIP pixels to LCTV pixels. The measured ratio agreed well with the value determined by Juday's method. Although it was not done because of time constraints, it is also possible to measure phase shift by this approach by measuring the beat position. Screen uniformity can be checked by looking for any bowing in the moiré beat.

The mapping experiments yielded the following affine (line) relation between PIP pixel coordinates and LCTV coordinates:

$$t = 1.522u + 21.621$$

where t and u are PIP and LCTV pixel coordinates respectively. The equation assumes that the first pixels in a row are at 0. It can be used to remap the array coordinates of the output from HOLOMED. to pixel coordinates in the PIP. Once this inverse mapping of the desired LCTV values is determined, it is easy to determine the corresponding gray levels at pixel coordinates within the PIP. These remapped coordinates are typically "virtual" since only integer coordinate values can be realized within the framegrabber. This requires that the remapped values be interpolated using to get the actual values written to the PIP. This was done using the Matlab's cubic spline fit function. The m-file for calculating the inverse mapping and performing the interpolation is LCTVFIT.M. This subprogram is called under the m-file TVPIP.M which takes the file output from HOLOMED as an argument and produces a file that can be written directly to the PIP provided the Matlab header is stripped.

OBSERVATION OF THE VIRTUAL IMAGE

The true test of all the laboratory measurements and the methodologies used to calculate fields and optimally realize them on the LCTV was the quality of the resulting image observed on the LCTV. The virtual image on the LCTV was calculated to be an F that was 6 mm high and one meter behind the LCTV. It was assumed that the hologram would be illuminated with a collimated beam from a helium-neon laser ($\lambda = .6328$ microns). A 50 cm lens was placed about 4.5" from the LCTV and used to reimage the F on a CCD television camera. An highly astigmatic image appeared about 32 inches from the lens. The calculated distance was about 35 inches. Multiple images of the F were observed, due to the discrete grating structure of the LCTV. The error in the image location and the astigmatism were eventually led to the discovery of an error in the computer code. The pixel pitches in the x and y directions had been entered in reverse order. This error leads to severe astigmatism and errors in the image position. After correcting the coding error the astigmatism was gone. The high quality image shown in Fig. 1 was produced. Note that the point sources that make up the 'F' are clearly separated and can be used to estimate the resolution of the hologram.

SUMMARY AND CONCLUSIONS

Methods of calculating and optimally realizing a hologram have been developed and proven in the laboratory. The following has been accomplished:

(1) Efficient code that uses an outer product formulation for calculating fields for objects consisting of elemental points and rectangular apertures has been developed.

(2) The MEDOF algorithm used for filter calculations has been successfully adapted to handle metrics appropriate for holography.

(3) Two laboratory methods have been developed for determining an accurate mapping of framegrabber pixels to LCTV pixels.

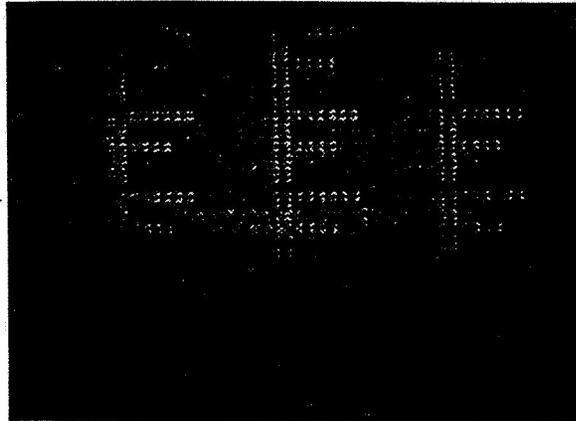


Figure 1. The image from the LCTV hologram of the letter F.

(4) A friendly software system has been developed that integrates the hologram calculation and realization process.

(5) The code and all the laboratory measurements on the SLM parameters has been proven with a test image.

FUTURE RESEARCH

A good deal of work is needed to completely exploit the hologram production system that has been developed. A possible (but far from exhaustive list) includes testing of

- (1). variations in mapping parameters
- (2). various coupling curves and their effects on images
- (3). of different hologram metrics
- (4). new SLMs
- (5). 3-D images
- (6). dynamic imagery

REFERENCES

1. C. Soutar, S. E. Monroe, Jr. and J. Knopp, "Measurement of the complex transmittance of the Epson liquid crystal television", *Optical Engineering*, **33**, 1061-1068 (1994).
2. W. H. Lee, "Sampled Fourier Transform Hologram Generated by Computer", *Appl. Opt.*, **9**, 639 (1970).
3. A. W. Lohman and D.P. Paris, "Computer Generated Spatial Filters for

Coherent Data Processing", *Appl. Opt.*, **7**, 651 (1968).

4. D. Gabor, "A New Microscopic Principle", *Nature*, **161**, 777, (1948).

5. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York (1968).

6. M. Abramowitz and I. Stegun (editors), *Handbook of Mathematical Formulas*, U.S. Department of Commerce, 7th Printing, page 300, May (1968).

7. R. D. Juday, "Optimal realizable filter minimum Euclidean distance principle", *Appl. Opt.*, **32**, 5100, (1993).

8. Details of this approach will be reported later in a paper by R.D. Juday and J. Knopp et. al. to be submitted to the *SPIE*. conference in April, Orlando, Fl. (1996) and eventually offered to *Optical Engineering*.

9. Matrox Inc., 1055 Regis Blvd., Dorval, Quebec, Canada, H9P2T4.

**INTELLIGENT PLANNING AND SCHEDULING FOR
CONTROLLED LIFE SUPPORT SYSTEMS**

**Final Report
NASA/ASEE Summer Faculty Fellowship Program 1995
Johnson Space Center**

Prepared by:	V. Jorge Leon, Ph.D.
Academic Rank:	Assistant Professor
University & Department	Texas A&M University Engineering Technology and Industrial Engineering College Station, TX 77843
NASA/JSC	
Directorate:	Engineering
Division:	Automation, Robotics & Simulation
Branch:	Intelligent Systems
JSC Colleague	Mariata McCamy
Date Submitted	August 11, 1995
Contract Number	NGT-44-001-800

ABSTRACT

Planning in Controlled Ecological Life Support Systems (CELSS) requires special lookahead capabilities due to the complex and long-term dynamic behavior of biological systems. This project characterizes the behavior of CELSS, identifies the requirements of intelligent planning systems for CELSS, proposes the decomposition of the planning task into short-term and long-term planning, and studies the crop scheduling problem as an initial approach to long-term planning.

CELSS is studied in the realm of Chaos. The amount of biomass in the system is modeled using a bounded quadratic iterator. The results suggests that closed ecological systems can exhibit periodic behavior when imposed external or artificial control.

The main characteristics of CELSS from the planning and scheduling perspective are discussed and requirements for planning systems are given. Crop scheduling problem is identified as an important component of the required long-term lookahead capabilities of a CELSS planner. The main characteristics of crop scheduling are described and a model is proposed to represent the problem. A surrogate measure of the probability of survival is developed. The measure reflects the absolute deviation of the vital reservoir levels from their nominal values. The solution space is generated using a probability distribution which captures both knowledge about the system and the current state of affairs at each decision epoch. This probability distribution is used in the context of an evolution paradigm.

The concepts developed serve as the basis for the development of a simple crop scheduling tool which is used to demonstrate its usefulness in the design and operation of CELSS.

INTRODUCTION

Significant automation will be required for the operation of Controlled Ecological Life Support Systems (CELSS) in order to enable the crew to spend more time carrying out science and mission related activities rather than routine, but indispensable, life support related activities. The successful control of a CELSS will depend to a great extent on our ability to predict the behavior of highly restrained biological systems and to maintain stable balance between the crew, biological, mechanical and physico-chemical systems.

Long-term dynamic and non-linear behavior characterize ecological systems. For instance, the sole decision of planting a given crop today must be associated with a series of activities and events which will both produce and consume vital resources for a period of time measurable in months or years. More over, the appropriateness of a planting decision depends on what the current state of affairs is - i.e., what other crops are currently in growth, what are the O₂, CO₂, and food storage levels, energy status, etc. Managing and controlling this type of system present a formidable task which may be impossible to deal with manually.

This project studies issues associated with intelligent planning and scheduling as means to aid in the design, operation and behavior prediction of CELSS. This report will present the planning and scheduling models and methodologies developed during the summer program. The implementation of a crop scheduling tool using these concepts is presented in a separated report in this volume (Whitaker and Leon, 1995).

Given a set of goals, a set of allowable actions, and an a description of an initial state of affairs, *planning* is defined as the task of finding a sequence of actions that will bring about a state of affairs in which all the desired goals are satisfied (Kautz and Pednault, 1988). *Scheduling* is defined here as the resolution of time conflicts generated by actions competing for scarce resources. It must be noted that, planning can also assign resources to actions; however, time conflicts on the usage of resources is only partially specified in the form of precedence relations between actions. On the other side, scheduling assumes that the sequence of actions required to accomplish a given goal is given in advance and only deals with the appropriate timing between these actions. It is well known that most practical cases of planning and scheduling problems are very complex problems proven to be mathematically intractable from the optimization or satisficing point of view. Figure 1 illustrates a sample plan generated to accomplish the goal "harvest wheat at time 1 and plant lettuce at time 2." The plan will also specify resource assignment to action. For instance, "crew person No.2" is assigned to perform the action "select seeds."

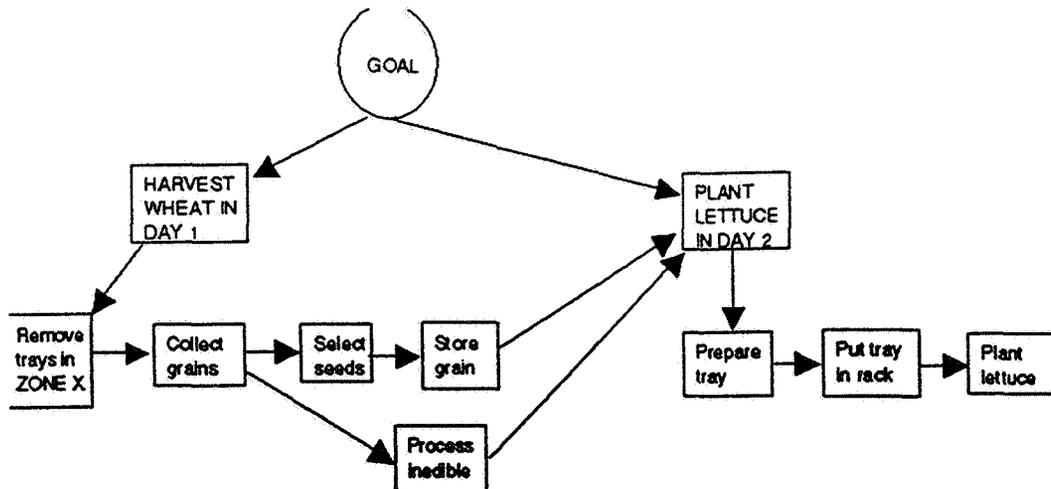


Figure 1. A typical plan

CHARACTERISTICS OF CELSS

From the planning, scheduling and control perspective, the inclusion of biological systems is what makes CELSS unique when compared to most systems studied in the literature. Species in biological systems adapt to changes in the environments using their internal mechanisms of control. In fact, it has been suggested that species may adopt the strategy of instability in order to enhance their chances of survival (Colombano, 1981). Thus, it may not be appropriate to equate stability with survival. Another complicating fact is that, biological systems can be controlled through the careful manipulation of external environmental variables; however, the relation between the control action and the systems's response can be very indirect, subject to extensive filtering and occasional reinterpretation (MacElroy, 1981). Simple CELSS models have been used to demonstrate that it is possible that failures can occur at times long after the cause of the perturbation has been removed (Auslander, 1981). For instance, it is possible that the percent of edible biomass obtained from a given crop can be severely decreased if at some point during growth the plants were subjected to long periods of darkness. In this example, the time elapsed between the disturbance (i.e., darkness) and the effect (i.e., harvest) can be in the order of months or several weeks.

The above characteristics make the design and operation of CELSS a formidable challenge which we are only starting to understand. Among the unanswered questions there are those related to the amount of local versus global control required, the quantification of the probability of survival, the sizing of vital element reservoirs, and

others. New paradigms may be necessary to model and analyze these systems. In this project, we explored two of a number of possible paradigms. Specifically, *Chaos* was used to model the non-linear and dynamic behavior of the system, and *Evolution* is used as a paradigm to generate a space of crop schedules.

CELSS - A Perspective From Chaos

A main difference between "natural" or "free" biological systems and CELSS resides in the latter's imposition of external bounds on the system's response. In this section we describe how Chaos theory can be used to study the response of biological systems (i.e., amount of biomass) when bounded externally. A simple experiment demonstrates that an unstable "free" biological system may have a periodic behavior when "constrained" externally.

Chaos theory has been recently used to model population dynamics and the behavior of biological and other natural systems (May, 1976). Chaotic behavior can be represented by simple mathematical models - however, the resulting behavior may be unpredictable. Chaos can be used to model the behavior of systems that are non-linear and sensitive to initial conditions. Non-linearity implies that what occurs now significantly affects future events. As with biological behavior, chaotic behavior is a collection of many orderly behaviors, none of which dominates each other under ordinary circumstances. The explicit consideration of this apparent instability in systems behavior may enable the development of better models for the analysis and synthesis of controlled systems. For instance, chaotic systems have been controlled by perturbing them in the right way so they will be forced to follow a different behavior. Furthermore, these controllers have proved to be more efficient than their traditional counterparts.

The quadratic iterator known as the Logistic Function (Velhurst, 1984) is used in the experiments. This function can be expressed as follows:

$$x(n+1) = a x(n) (1 - x(n))$$

Where, $x(n)$ is the normalized size of the population of a species at time n , and a is a proportionality constant. This function has interesting characteristics, such as:

1. The size of the population at any time depends on the initial condition ($x(0)$).
2. Stable, periodic or unstable behavior can be represented with the appropriate choice of the parameter a .
3. It can portray abrupt changes in behavior from order to chaos - i.e., period-doubling bifurcations.
4. It can portray long-term stability through "attracting" states.
5. It exhibits "universal" behavior observed in many natural systems.

These characteristics make the apparently random behavior of the systems predictable to some extent. The important result is that this predictability allows for the control of chaotic systems; in fact, there is evidence that control using Chaos can outperform traditional control (Ditto and Pecora, 1993).

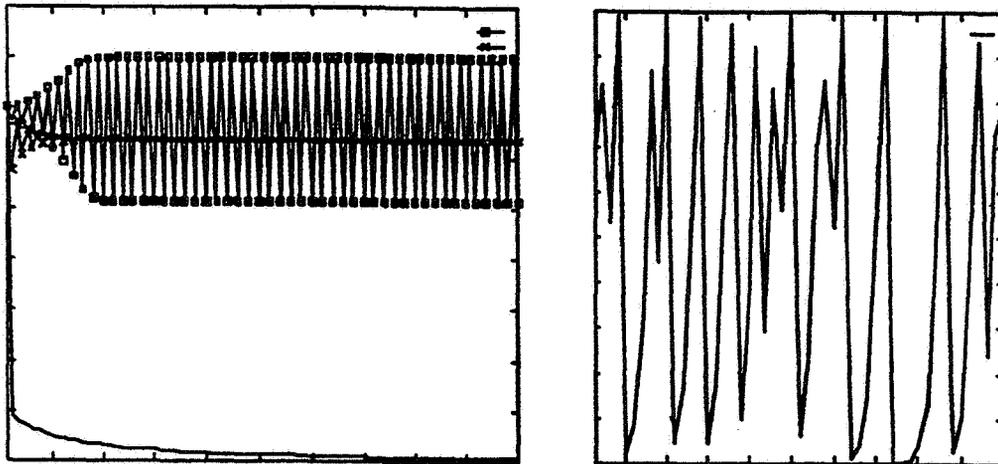
Figure 2 (a), (b) and (c) show plots of the same function for different values of a resulting in stable, periodic and unstable behavior, respectively.

We model a CELSS using the logistic function; where, $x(n)$ represents the amount of plant biomass in the system. Clearly, in a CELSS this amount cannot grow unbounded. In order to model this, an upper bound, U , is imposed limiting the maximum amount of biomass at any point in time. The function used in the experiment is modified as follows:

$$x(n+1) = a x(n) (1 - x(n))$$

if $x(n+1) > U$, then set $x(n+1) = U$.

The value of $a=4.0$ applied to the original "free" system yields the unstable behavior depicted in Figure 2(b). However, the behavior of the "bounded" system became periodic as illustrated in Figure 3.



(a) Extinction, stable and periodic

(c) Unstable, $a = 4.0$

Figure 2 .The Logistic Function

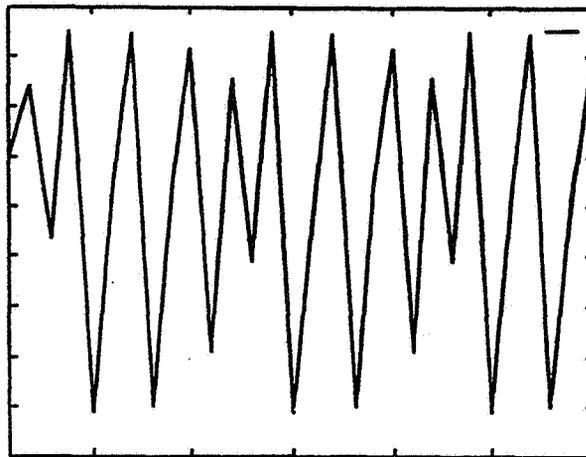


Figure 3. A bounded Logistic function with $a = 4.0$, $U=0.9$

This interesting result using a simple experiment suggests that the implementation of a CELSS may be possible. Further, it suggests that the system will have alternating periods of time in which the biological system is mostly controlled using the plant's internal control mechanisms and period during which artificial control will be needed. For instance, incineration or other recycling process may be required whenever the amount of food produced is excessive.

This simple experiment also suggests that more formal and thorough studies using the Chaos paradigm may be valuable for the understanding of CELSS.

CELSS PLANNERS - REQUIREMENTS

The long duration of CELSS operations and the complex dynamics induced by the biological systems, mechanical reliability and information uncertainty make it necessary for a CELSS planner to have the following special requirements:

1. Planners must explicitly consider the long-term dynamics (i.e., in the order of several months) inherent to biological systems. This implies the need of lookahead capabilities not present in traditional planners. The lookahead should ensure long-term system stability and strategic management of resources. Only vital resources must be considered and time granularity should be in the order of days.

2. Detail resource scheduling and minute-to-minute planning should be done using a planning horizon in the order of hours to days. It is important that detailed plans are consistent with the long-term lookahead analysis.

3. Planners must exploit the intelligence and flexibility of humans. This implies that planners should: (a) enable the incorporation of user input before and during plan generation, (b) provide with plan explanation and (c) provide with "what-if" analysis capabilities.

4. Planners must be adaptable to unforeseen changes in operating conditions. CELSS may need to operate in isolation from terrestrial feedback for months. The planner must be able to accommodate to unknown or unforeseen conditions during the design of the system. Planners must allow the incorporation of experience for adaptation to new conditions and improvement in performance.

PLANNING PROBLEM DECOMPOSITION

The requirements described above suggest the decomposition of the planning problem into two subproblem: Long-Term Planning (LTP) and Short-Term Planning (STP). Figure 4 depicts the proposed problem decomposition. STP is much like most planning problems in the literature; i.e., is well described by the definition of planning in the beginning of this report. Although very important, this will not be discussed further in this report.

The LTP can be described as the problem of scheduling strategic resources. Strategic resources may include crew, crops, main elements (O₂, CO₂, H₂O, etc.), food, storage space and others. Figure 4 shows how only the initial portion of a long-term plan is broken down into further detail in the short-term plan. Further, it

also suggests that the immediate actions prescribed by LTP ensure the adequate levels in the long run. It must be noted that LTP gives the planning tool with predictive capabilities which make it also useful as a "what-if" tool for design and situation diagnosis.

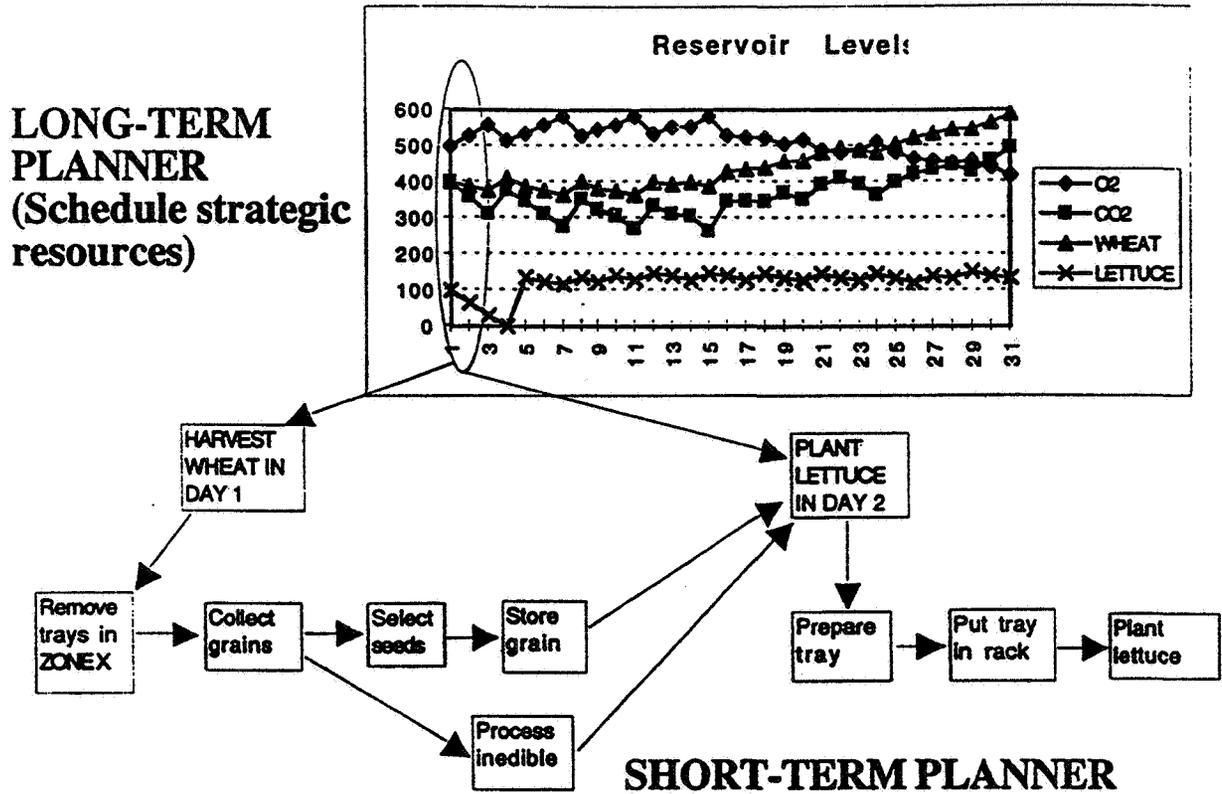


Figure 4. Planning problem decomposition

THE CROP SCHEDULING PROBLEM

The crew, crops and other live agents determine to a great extent the long-term behavior of a CELSS. In this project, a first look is given to the Crop Scheduling Problem (CSP). CSP can be formulated as follows:

CSP:

- Decision Variables: What, how much and when to plant.
- Objective: Maximize the probability of survival.
- Constraints: Conservation and transformation of mass
Crew availability
Space availability (planting, process, storage)
Energy and other operational constraints

This scheduling problem is unique for a number of reasons which make it specially challenging when compared with more traditional scheduling problems. The first distinction is related to the time characteristics of plants as illustrated in Figure 5. Once a decision of planting is made, this triggers a several events in the near (same day) and far (months later) future. Although some precedence exists between these events, some partial sequences may occur simultaneously and may require of the same scarce resources. The management of such a system becomes too complicated to deal with manually, specially if different types of crops are considered. For instance in Figure 5, at the time of planting one must consider how will this affect the O_2 and CO_2 reservoirs during its growth and how will it impact the food storage space, crew loading and waste processing about 3 months later (e.g., wheat). The non-linearity of the problem becomes evident when observing that other planting decisions will have to be made while the events triggered by the decision of planting are still taking place.

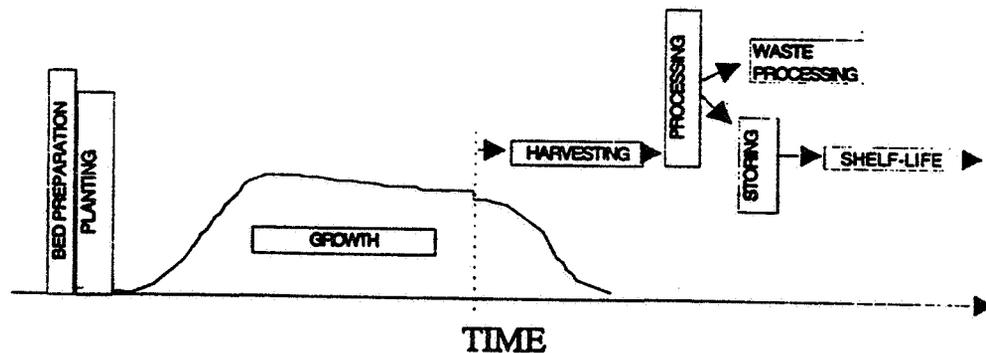


Figure 5. A plant model

A CROP SCHEDULING MODEL

In this section a model for CSP is proposed. First, a surrogate measure of the probability of survival is proposed as an objective function. Second, a solution space that is generated using an evolution paradigm is described. Finally, the search and ideas to deal with the size of the solution space will be discussed.

Objective Function

The quantification of the probability of survival is not a trivial issue. This problem is similar to the one in safety analysis but in CELSS is complicated by the long-term dynamics inherent to such systems as growing plants (Auslander, 1981). Here, the probability of survival is represented using a surrogate measure, Z , which reflects the absolute deviations of the each vital reservoir's level from the corresponding nominal value.

Let d_r be some normalized indicator (i.e., between 0 and 1) of the average deviation of reservoir's r level from its nominal value. Z can be defined as

$$Z = f(d_1, d_2, \dots, d_R)$$

where, R is the total number of vital reservoirs and $f(.)$ is an appropriate real valued function.

Solution Space

The state of the system is defined by the status of the in-growth crops, number of empty trays and reservoir levels. Other important information are the time at which the state is described, crew profile, human model, plant models and physical system models.

At a decision epoch, a decision is made as for what, how much and when to plant. Decision epochs are prompted by significant events. These events can be classified into (i) biological events, (ii) user specified events and (iii) stochastic events.

The solution space can be represented as a tree where the nodes are decision points and the branches are different paths resulting from different scheduling decisions. A path in this tree will represent one possible schedule. A schedule which satisfies all the operating constraints is termed an admissible plan. The scheduling problem can be stated as the problem of finding the admissible plan which minimizes Z .

Considering all possible alternatives at each decision epoch would be impractical. Rather, we suggest the generation of alternatives using heuristics which capture both knowledge about the problem, as well as, the current state of affairs. Inspired by evolution processes, a fitness probability distribution is determined at each decision epoch. The random variable is the crop type, and the corresponding probability reflects the marginal effect that planting the corresponding crop will have on the vital reservoirs. These probabilities can be computed using arguments similar to the ones used in the determination of the objective function. Knowledge about the process is captured through the use of the system's models to predict the impact that each crop would have if planted. The current state of affairs is captured since the reservoir levels at the decision epoch must be considered. It must be noted that this probability distribution must be computed at each decision epoch to reflect the updated "desirability" of each crop. Thus the term *desirability probability*.

There are several different ways that this probability can be used to generate the solution space. The most efficient way of using it is still a matter of further research. One way to implement this generation strategy is illustrated in Figure 6. Given a number of empty trays, one can randomly sample from the desirability distribution until all empty trays are filled - noted that, *plant-nothing* is considered as one possible crop. Figure 6 illustrate how a single path (schedule) can be generated.

Search and Dealing With Complexity

Clearly, generating the space using local information will unavoidably lead to the necessity of backtracking when non-admissible situations are encountered. Too

much backtracking may render the approach impractical. In the case of CELSS, however, too much backtracking may suggest that the system is not robust enough; i.e., there are only a few paths leading to mission completion. If this is the case, a system redesign will be more recommendable than a more sophisticated scheduler. Clearly, too much backtracking may also suggest a poor space generation scheme.

For CELSS it would be desirable to allow for user intervention if conflicts cannot be resolved automatically. Hence, the importance of schedule explanation to aid the user in suggesting conflict resolution.

Most existing search strategies may be applied to deal with the size of the solution space. The determination of the most appropriate search strategy is still an open research issue.

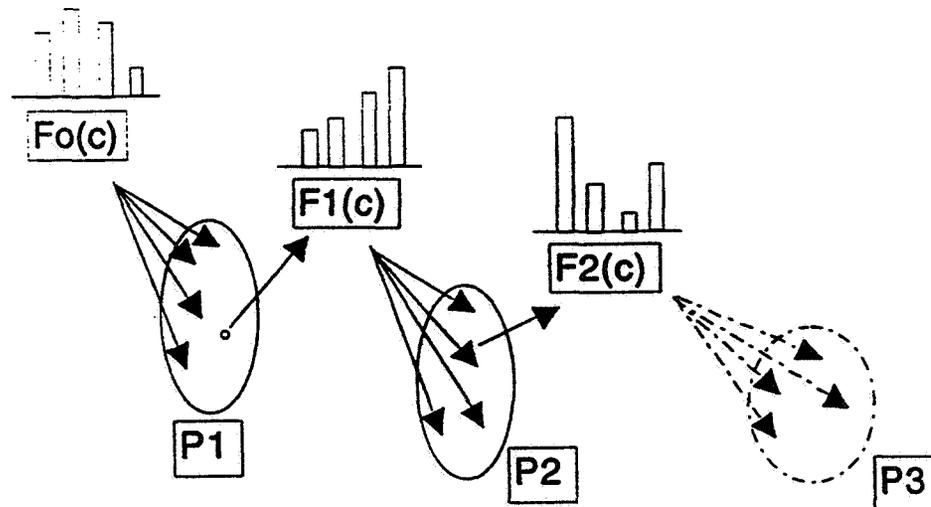


Figure 6. An evolution approach

AN EXAMPLE IMPLEMENTATION OF THE A SCHEDULER

A simple crop scheduler was implemented to illustrate its potential use in planning as well as design of CELSS. A detailed description of this implementation is contained in a separate report (Whitaker and Leon, 1995). Figure 7 illustrate the main components of the Intelligent Crop Scheduler (ICS) developed during this summer project. ICS has two main components: a Schedule Generator (SG) and a

Schedule Simulator (SS). SG is based on the concepts described above. SS contains detail human metabolic simulator and two plant models (i.e., wheat and lettuce). It also contains a simplified physical system model as illustrated in Figure 8. For simplicity, only O₂, CO₂, edible wheat and edible lettuce reservoirs are considered. The main systems considered are the crew, plants and a generic waste processing system.

A sample planting schedule is illustrated in Figure 9. Figure 10 illustrates the reservoir levels for the schedule, as well as, its performance. A variety a scenarios were run illustrating how the output can be used to aid in sizing the gas tanks, food storage space, growing area, crew profile, planting strategy, growing parameters, and others. See Whitaker and Leon (1995) for a discussion of the example cases.

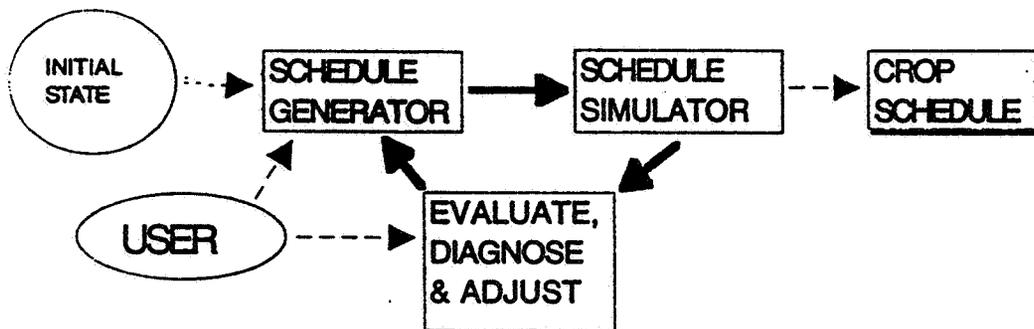


Figure 7. Intelligent Crop Scheduler

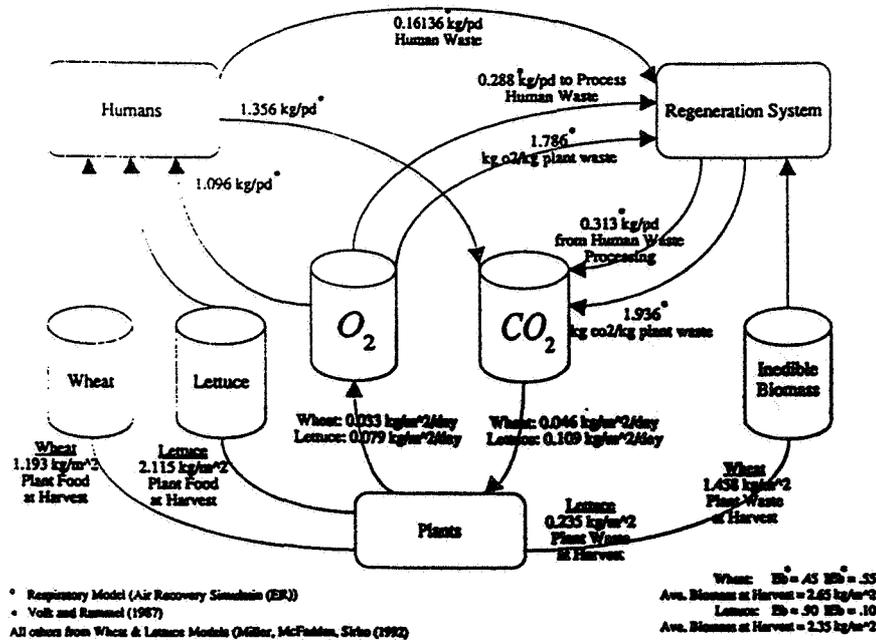


Figure 8. A simplified CELSS model

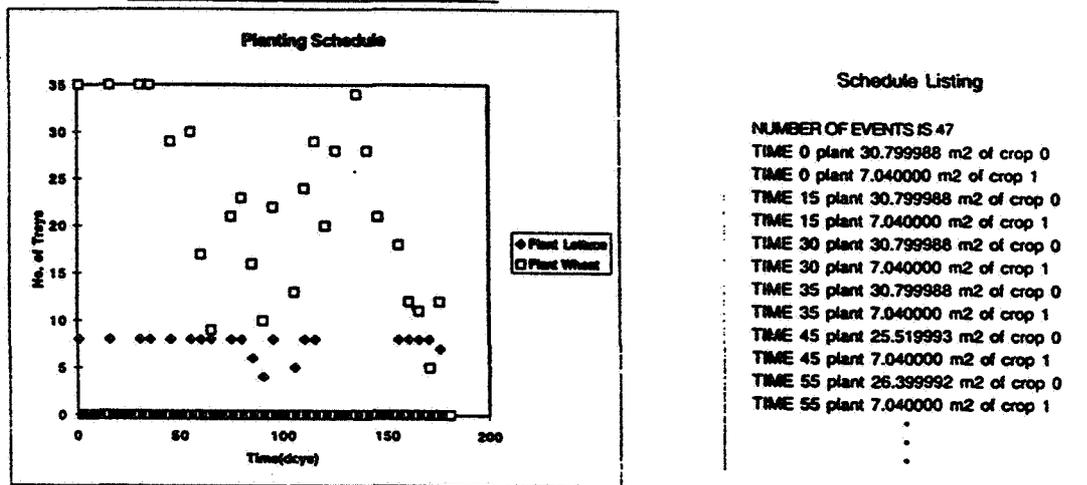


Figure 9. Sample crop schedule

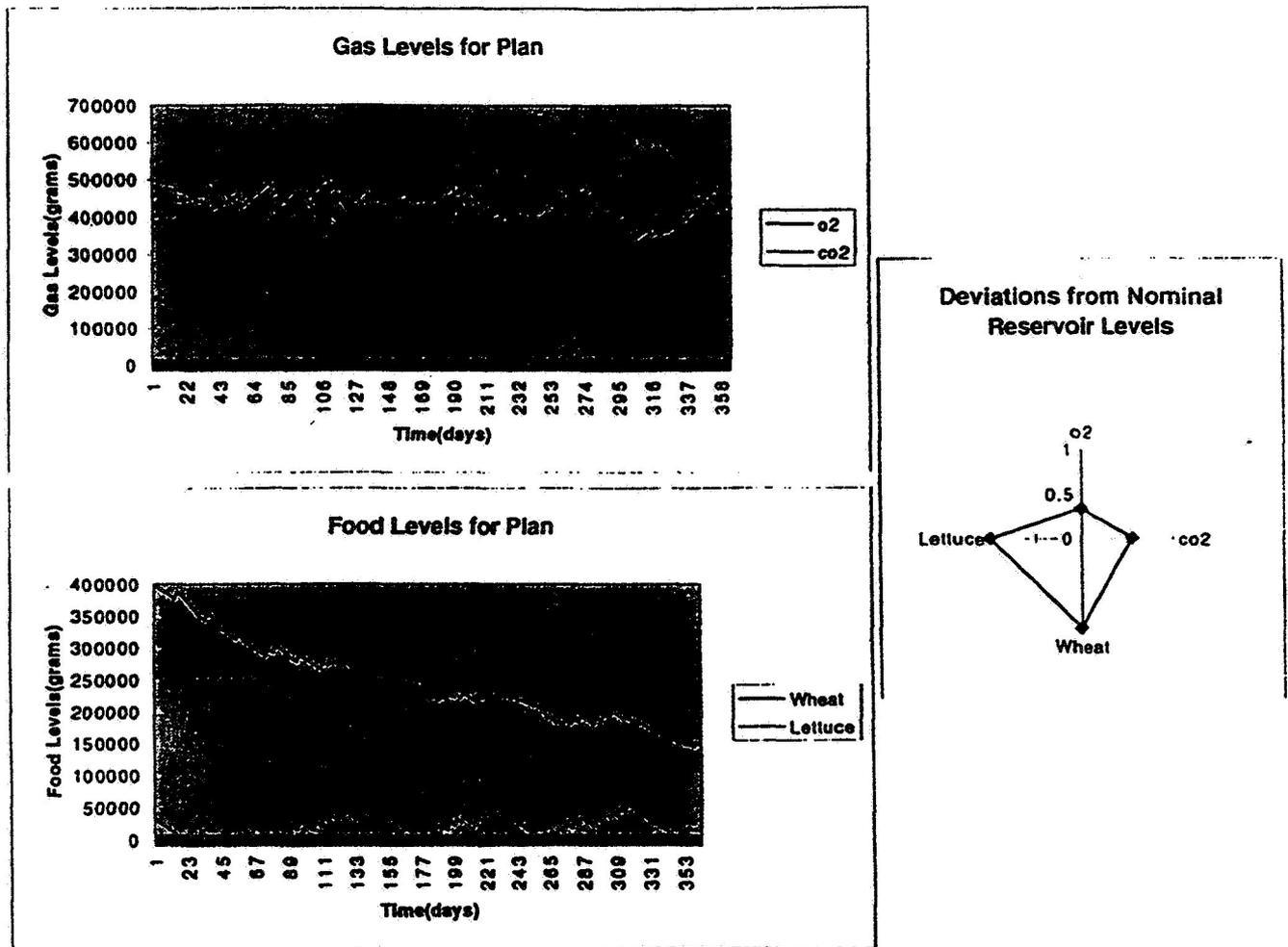


Figure 10. Sample reservoir level output

REFERENCES

Ditto, W.L. and L.M. Pecora (1993) "Mastering Chaos," *Scientific American*, August 1993.

Gelick, J. (1987) *Chaos - Making a New Science*, Viking.

Kautz, H.A. and E.P.D. Pednault (1988) "Planning and Plan Recognition," *AT&T Technical Journal*, 67, 25-40.

May, R.M. (1976) "Simple Mathematical Models With Very Complicated Dynamics," *Nature*, 261, 459-467.

The articles for references to MacElroy (1981), Auslander (1981) and Colombano (1981) are in NASA Publication 2247: "Controlled Ecological Life Support System - First Principal Investigators Meeting," *Proceedings of a NASA Meeting*, University of New Hampshire, Durham, NH, May 3-6, 1981.

Peitgen, H.-O., H. Jurgens and D. Saupe (1992), *Chaos and Fractals - New Frontiers of Science*, Springer-Verlag, New York, 1992.

Volk, T. and J.D. Rummel (1987) "Mass Balances for a Biological Life Support System Simulation Model," *Adv. Space Res.*, 7 (4), 141-148.

Whitaker, L.O. and V.J. Leon (1995) "An Intelligent Crop Scheduling Tool for CELSS," 1995 NASA/ASEE Summer Faculty Fellowship Program, NASA - Johnson Space Center, Houston, Texas.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE August 1996	3. REPORT TYPE AND DATES COVERED NASA Contractor Report		
4. TITLE AND SUBTITLE National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program - 1995 (Volumes 1 and 2)			5. FUNDING NUMBERS NGT 44-001-800	
6. AUTHOR(S) William A. Hyman* and Donn G. Sickorez, Editors				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Lyndon B. Johnson Space Center Office of Public Affairs Houston, Texas 77058			8. PERFORMING ORGANIZATION REPORT NUMBERS	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, D. C. 20546-0001			10. SPONSORING/MONITORING AGENCY REPORT NUMBER CR-201377	
11. SUPPLEMENTARY NOTES *Texas A&M University, College Station, Texas				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified/Unlimited Available from the NASA Center for AeroSpace Information (CASI) 800 Elkridge Landing Road Linthicum Heights, MD 21090-2934 (301) 621-0390			12b. DISTRIBUTION CODE Subject Category: 99	
13. ABSTRACT (Maximum 200 words) The JSC NASA/ASEE Summer Faculty Fellowship Program was conducted at JSC, including the White Sands Test Facility, by Texas A&M University and JSC. The objectives of the program, which began nationally in 1964 and at JSC in 1965, are (1) to further the professional knowledge of qualified engineering and science faculty members; (2) to stimulate an exchange of ideas between participants and NASA; (3) to enrich and refresh the research and teaching activities of the participants' institutions; and (4) to contribute to the research objectives of the NASA centers. Each faculty fellow spent at least 10 weeks at JSC engaged in a research project in collaboration with a NASA/JSC colleague. In addition to the faculty participants, the 1995 program included five students. This document is a compilation of the final reports on the research projects completed by the faculty fellows and visiting students during the summer of 1995. The reports of two of the students are integral with that of the respective fellow. Three students wrote separate reports.				
14. SUBJECT TERMS information transfer, research, research projects, engineering, science, university program			15. NUMBER OF PAGES 428	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	